

Skill of seasonal flow forecasts at catchment scale: an assessment across South Korea

Yongshin Lee¹, Francesca Pianosi¹, Andres Peñuela², Miguel Angel Rico-Ramirez¹

¹ School of Civil, Aerospace and Design Engineering, University of Bristol, Bristol, BS8 1TR, UK

² Department of Agronomy, Unidad de Excelencia María de Maeztu, University of Cordoba, Cordoba, 14071, Spain

Correspondence: Yongshin Lee (yongshin.lee@bristol.ac.uk)

Abstract.

Recent advancements in numerical weather predictions have improved forecasting performance at longer lead times. Seasonal weather forecasts, providing predictions of weather variables for the next several months, have gained significant attention from researchers due to their potential benefits for water resources management. Many efforts have been made to generate Seasonal Flow Forecasts (SFFs) by combining seasonal weather forecasts and hydrological models. However, producing SFFs with good skill at a finer catchment scale remains challenging, hindering their practical application and adoption by water managers. Consequently, water management decisions, both in South Korea and numerous other countries, continue to rely on worst-case scenarios and the conventional Ensemble Streamflow Prediction (ESP) method.

This study investigates the potential of SFFs in South Korea at the catchment scale, examining 12 reservoir catchments of varying sizes (ranging from 59 to 6648 km²) over the last decade (2011-2020). Seasonal weather forecasts data (including precipitation, temperature and evapotranspiration) from the European Centre for Medium-Range Weather Forecasts (ECMWF system5) is used to drive the Tank model (conceptual hydrological model) for generating the flow ensemble forecasts. We assess the contribution of each weather variable to the performance of flow forecasting by isolating individual variables. In addition, we quantitatively evaluate the overall skill of SFFs, representing the probability of outperforming the benchmark (ESP), using the Continuous Ranked Probability Skill Score (CRPSS). Our results highlight that precipitation is the most important variable in determining the performance of SFFs, and temperature also plays a key role during the dry season in snow-affected catchments. Given the coarse resolution of seasonal weather forecasts, a linear scaling method to adjust the forecasts is applied, and it is found that bias correction is highly effective in enhancing the overall skill. Furthermore, bias corrected SFFs have skill with respect to ESP up to 3 months ahead, this being particularly evident during abnormally dry years. To facilitate future applications in other regions, the code developed for this analysis has been made available as an open-source Python package.

Keywords: Seasonal weather forecasts, Seasonal flow forecasts, Skill assessment, Ensemble Streamflow Forecast, CRPSS, Linear scaling

1. Introduction

Over the last decade, numerical weather prediction systems have improved their forecasting performance at longer lead times, ranging from 1 to several months ahead (Alley et al., 2019; Bauer et al., 2015). The water management sector may benefit considerably from these advances. In particular, predictions of weather variables such as precipitation and temperature several months ahead ('seasonal weather forecasts' from now on) might be exploited to anticipate upcoming dry periods and implement management strategies for mitigating future water supply deficits (Soares and Dessai, 2016).

To increase relevance for water resource management, seasonal weather forecasts can be translated into Seasonal Flow Forecasts (SFFs) via a hydrological model. SFFs can be provided and evaluated at different temporal and spatial resolutions: a coarser resolution, e.g., magnitude of total next-month runoff over a certain region (Arnal et al., 2018; Prudhomme et al., 2017) or a finer resolution, e.g., daily/weekly flow at a particular river section over the next month (Crochemore et al., 2016; Lucatero et al., 2018). This distinction is important here because coarser resolution SFFs can only be applied to inform water management in a qualitative way, whereas finer resolution SFFs can also be used to force a water resource system model for a quantitative appraisal of different management strategies. Proof-of-principle examples of the latter approach are provided by Boucher et al. (2012), Chiew et al. (2003), and Peñuela et al. (2020). These papers have demonstrated, through model simulations, the potential of using SFFs to improve the operation of supply reservoirs (Peñuela et al., 2020), irrigation systems (Chiew et al., 2003) and hydropower systems (Boucher et al., 2012).

54 Obviously, generating SFFs with good skill at finer scales is challenging and the lack of forecasting performance
55 is often cited as a key barrier to real-world application of SFFs by water managers (Jackson-Blake et al., 2022;
56 Soares and Dessai, 2016; Whateley et al., 2015). In practice, if a Water Resource System (WRS) model is used to
57 simulate and compare different operational decisions, this is done by forcing the WRS model against a repeat of
58 a historical low flow event (“worst-case” scenario) (Yoe, 2019) or against the Ensemble Streamflow Prediction
59 (ESP). ESP is a widely used operational forecasting method whereby an ensemble of flow forecasts is generated
60 by forcing a hydrological model with historical meteorological observations (Baker et al., 2021; Day, 1985). Since
61 the hydrological model is initialised at current hydrological conditions, ESP is expected to have a certain level of
62 performance, particularly in ‘long-memory’ systems where the impact of initial conditions last over long time
63 periods (Li et al., 2009). Previous simulation studies that examined the use of SFFs to enhance the operation of
64 water resources systems (e.g., Peñuela et al., 2020, as cited above) did indeed show that ESP serves as a ‘hard-to-
65 beat’ benchmark. Similar to other countries, in South Korea, the worst-case scenario and ESP are used for
66 informing water management activities, whereas SFFs are not currently applied. Before the use of SFFs can be
67 proposed to practitioners, it is thus crucial to understand the skill of such products with respect to ESP.

68 Numerous studies were conducted on the skill of SFFs in different regions of the world. Some of these studies
69 focused on the ‘theoretical skill’, which is determined by comparing SFFs with pseudo-observations produced by
70 the same hydrological model when forced with observed temperature and precipitation. This experimental set-up
71 enables to isolate the contribution of the weather forecast skill to the flow forecast skill, regardless of structural
72 errors that may be present in the hydrological model. In general, most studies have found that the theoretical skill
73 of SFFs may be only marginally better than that of ESP in specific region and lead time. For example, Yoseff et
74 al. (2013) analysed multiple large river basins worldwide and found that SFFs generally perform worse than ESP.
75 Likewise, the findings of Greuell et al. (2019) indicated that SFFs are more skillful than ESP for the first lead
76 month only. Across Europe, the theoretical skill of SFFs was found to be higher than ESP in coastal and
77 mountainous regions (Greuell et al., 2018).

78 Although important to how the information content of seasonal weather forecasts vary across regions with
79 different climatic characteristics, from a water management perspective, the theoretical skill may not be the most
80 appropriate metric, as it reflects the performance within the modelled environment (Pechlivanidis et al., 2020)
81 rather than the real-world. The ‘actual skill’, which is determined by comparing SFFs to flow observations, would
82 be more informative for water managers to decide on whether to use SFFs and when. Previous studies that
83 investigated the actual skill showed that, as expected, the actual skill is lower than the theoretical skill due to
84 errors in the hydrological model and in the weather input observations (Greuell et al., 2018; van Dijk et al., 2013).
85 In addition, due to the coarse horizontal resolution of seasonal weather forecasts (around $1^{\circ} \times 1^{\circ}$), the forecast skill
86 can be significantly improved through bias correction, particularly of precipitation forecasts (e.g., Crochemore et
87 al., 2016; Lucatero et al., 2018; Tian et al., 2018). However, even after bias correction, SFFs were found unable
88 to surpass ESP in many previous applications (e.g., Crochemore et al., 2016; Greuell et al., 2019; Lucatero et al.,
89 2018).

90 Previous studies reviewed above have mainly used the seasonal weather forecasts provided by the European
91 Centre for Medium-Range Weather Forecasts (ECMWF). Here, it is important to note that the majority of these
92 studies have utilized ECMWF’s system 3 (e.g., Yossef et al., 2013) or 4 (e.g., Crochemore et al., 2016; Greuell et
93 al., 2019; Lucatero et al., 2018; Tian et al., 2018). A few studies comparing the performance of SFFs and ESP
94 have been conducted based on ECMWF’s cutting-edge forecasting system 5, which became operational in
95 November 2017. These include Peñuela et al., 2020 and Ratri et al., 2023, which however did not analyse the skill
96 of SFFs in much detail but rather focused on their operational implementation. Given that the upgrade of
97 forecasting system can lead to substantial enhancement in the performance (e.g., Johnson et al., 2019; Köhn-Reich
98 and Bürger, 2019), it is interesting to assess whether improved skill of weather forecasts delivered by the System
99 5 translates into improved skill of flow forecasts.

100 Our previous research (Lee et al., 2023) on the skill of seasonal precipitation forecasts across South Korea showed
101 that, among various forecasting centres, ECMWF provides the most skilful seasonal precipitation forecasts,
102 outperforming the climatology (based on historical precipitation observations). This is particularly evident during
103 the wet season (June to September) and in dry years, where skill can also be high at longer lead times beyond the
104 first month. Given the significant correlation between precipitation and flow in the country (Ministry of land,
105 infrastructure, and transportation, 2016), South Korea is an interesting test bed to investigate if the skill of seasonal
106 precipitation forecasts is mirrored into the skill of flow forecasts at the catchment scale.

107 Specifically, in this study we focus on 12 catchments of various size (from 59 to 6648 km²) which include the
108 most important multipurpose reservoirs across South Korea, and where the use of SFFs may be considered for
109 assisting operational decisions and mitigating impacts of droughts. Given this practical long-term goal, our study
110 focuses on assessing the ‘overall skill’, which represents the probability that SFFs outperform the benchmark

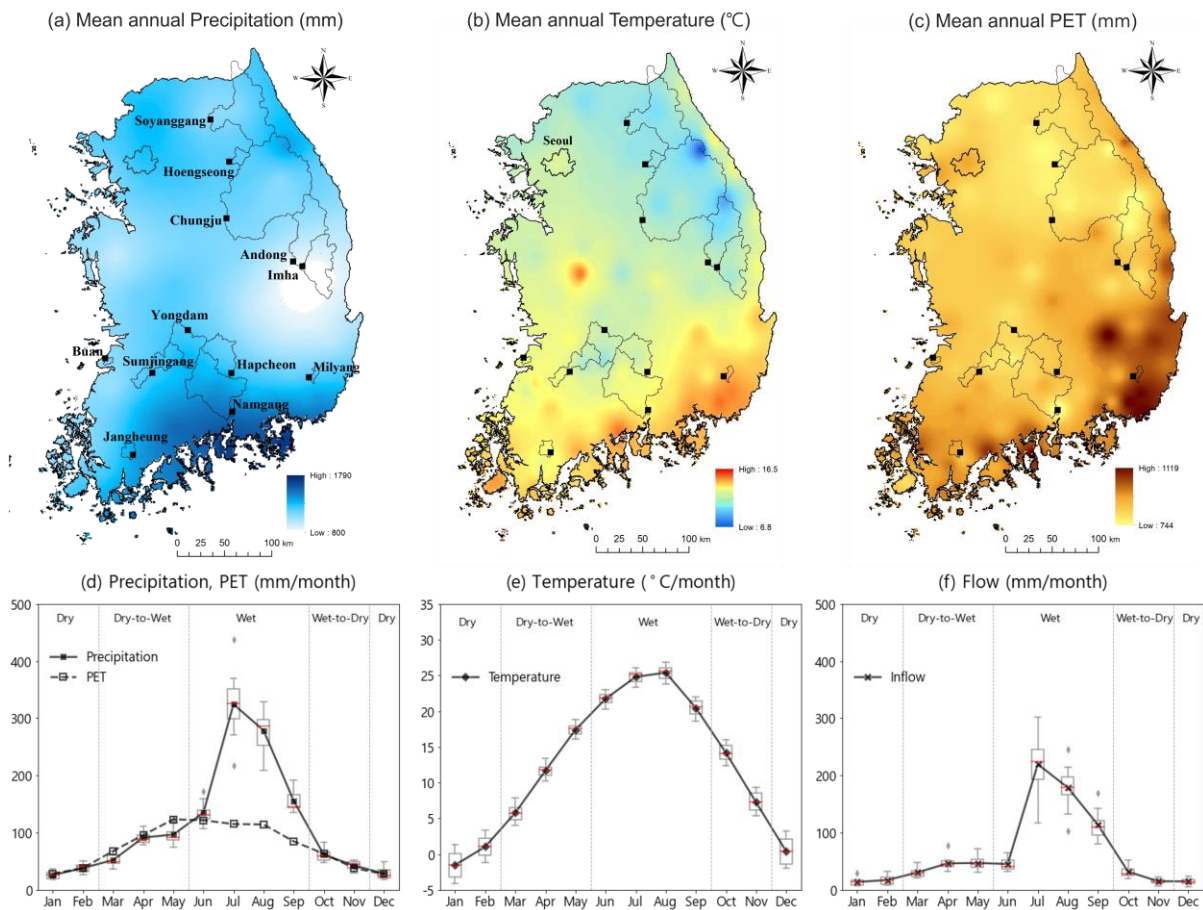
111 (ESP) when comparing the flow forecasts with historical flow observations. As a hydrological model, we use the
 112 lumped Tank model (Sugawara et al., 1986) which is the rainfall-runoff model currently in use for the national
 113 water management and planning. For all catchments, we briefly analyse the hydrological model performance, also
 114 investigate which weather forcing input (precipitation, temperature, and potential evapotranspiration) contributes
 115 most to the performance of SFFs across different catchments, before and after bias correction. Finally, we look at
 116 how the overall skill varies across seasons, years, and catchment, to draw conclusions on when and where SFFs
 117 may be more informative than ESP for practical water resources management. In doing so, we develop a workflow
 118 for SFFs analysis implemented in a Python Jupyter Notebook, which can be utilized by other researchers for
 119 evaluating and testing SFFs in various regions.

120 **2. Material and methodology**

121 **2.1 Study site and data**

122 **2.1.1 Study site**

123 The spatial scope of this study is defined as 12 multi-purpose reservoir catchments across South Korea. While
 124 there are 20 multi-purpose reservoirs nationwide (K-water, 2022), we have specifically selected 12 reservoirs with
 125 at least 10 years of flow observation and no external flows from other rivers or reservoirs. The location of the
 126 catchments and the mean annual precipitation, temperature, and potential evapotranspiration (PET) are shown in
 127 Figure 1(a-c). The weather data for the selected reservoir catchments is reported in Table 1.



128
 129 **Figure 1: Top row: mean annual (a) precipitation, (b) temperature and (c) PET across South Korea over the period**
 130 **1967-2020. Black lines are the boundaries of the 12 reservoir catchments analysed in this study (all maps obtained by**
 131 **interpolating point measurements using the inverse distance weighting method). Bottom row: (d) cumulative monthly**
 132 **precipitation and PET, (e) mean monthly temperature and (f) cumulative monthly flow. These three variables are**
 133 **averaged over the 12 reservoir catchments from 2001 to 2020. Box plots show the inter-catchment variability.**

134 **Table 1. Characteristics of the 12 multipurpose reservoirs (from North to South) and the catchments they drain (K-**
 135 **water, 2022). Tmin and Tmax represent mean monthly minimum and maximum temperature averaged over 2001-2020,**

136 all other meteorological variables (P: precipitation, T: temperature, PET: potential evapotranspiration) are annual
 137 averages over the same period.

Catchment	Soyanggang	Hoengseong	Chungju	Andong	Imha	Yongdam	Buan	Sumjingang	Hapcheon	Milyang	Namgang	Jangheung	
Area (km ²)	2703	209	6648	1584	1361	930	59	763	925	95	2285	193	
P (mm)	1220	1336	1197	1079	956	1317	1292	1343	1279	1375	1477	1439	
T (°C)	10.8	10.9	11.1	11.1	12.2	11.8	13.5	12.6	12.8	14.2	13.5	13.8	
Mean annual	T min	-4.2 (Jan.)	-4.0 (Jan.)	-3.2 (Jan.)	-3.5 (Jan.)	-1.6 (Jan.)	-2.3 (Jan.)	-0.1 (Jan.)	-1.5 (Jan.)	-0.8 (Jan.)	1.0 (Jan.)	0.4 (Jan.)	1.3 (Jan.)
	T max	24.0 (Aug.)	24.1 (Aug.)	25.9 (Aug.)	23.8 (Aug.)	25.1 (Aug.)	24.8 (Aug.)	26.7 (Aug.)	25.8 (Aug.)	25.5 (Aug.)	26.8 (Aug.)	26.0 (Aug.)	26.2 (Aug.)
PET (mm)	874	870	881	896	947	884	960	919	933	993	952	896	

138 Figure 1(d-f) shows the monthly changes in precipitation and PET (d), temperature (e) and flow (e) averaged over
 139 the 12 selected catchments from 2001 to 2020. Generally, the catchments located in the Southern region exhibit
 140 higher mean annual precipitation, temperature, and PET. In order to examine how the skill of seasonal weather
 141 and flow forecasts varies across a year, we divide a year into four seasons based on monthly precipitation (Lee et
 142 al., 2023): dry (December to February), dry-to-wet transition (March to May), wet (June to September), wet-to-
 143 dry transition (October to November). As shown in this figure, most of the total annual precipitation (and the
 144 corresponding flow) occurs during the hot and humid wet season, while the dry season is characterized by cold
 145 and dry conditions. The high inter-annual variability of precipitation and flow is a feature of South Korea's climate
 146 and is attributed to the impacts of typhoons and monsoons (Lee et al., 2023). Figure 1(d-f) also shows high inter-
 147 catchment variability during the wet season in both precipitation (d) and flow (f), whereas the inter-catchment
 148 variability in temperature (e) is more obvious during the dry season.

149 2.1.2 Hydrologic data and seasonal weather forecasts

150 Precipitation, temperature, and potential evapotranspiration are the key variables required to simulate flows using
 151 a hydrological model. To this end, daily precipitation data from 1318 in-situ stations produced by the Ministry of
 152 Environment, the Korea Meteorologic Administration (KMA), and the national water resources agency (K-water)
 153 (Ministry of Environment, 2021), and daily temperature data from 683 in-situ stations generated by KMA were
 154 obtained. Both precipitation and temperature data cover the period from 1967 to 2020 (see Figure 1). Potential
 155 evapotranspiration (PET) data was computed using the standardized Penman-Monteith method suggested by UN
 156 Food and Agriculture Organization. The precipitation and temperature measurements have been quality-controlled
 157 by the Ministry of Environment. We used the Thiessen polygon method to calculate the catchment average
 158 precipitation and temperature.

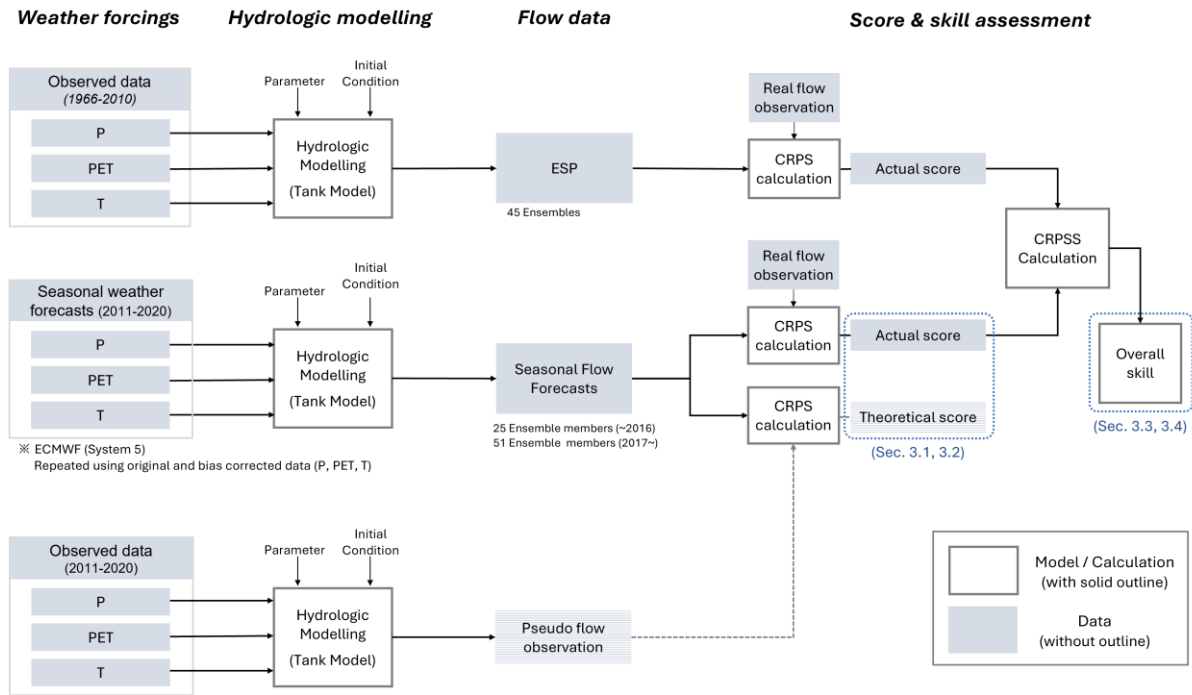
159 The flow data used in this study refers to the flow into the reservoir from their upstream catchment (see Table1
 160 and Figure 1). K-water generates daily flow data through a water balance equation, which takes into account the
 161 daily changes in reservoir volume (from storage-elevation curve) caused by the water level fluctuations and water
 162 supplies. However, to date, reservoir evaporation has not been considered in the flow estimation process. In this
 163 study, quality-controlled daily flow data for each reservoir produced by K-water is used.

164 Several weather forecasting centres, including ECMWF, the UK Met Office and the German Weather Service,
 165 provide seasonal weather forecasts datasets through the Copernicus Climate Data Store (CCDS). According to
 166 our previous study (Lee et al., 2023), ECMWF was found to be the most skilful provider of seasonal precipitation
 167 forecasts for South Korea. Since the precipitation is one of the most important weather forcings in hydrological
 168 forecasting (Kolachian and Saghafian, 2019), we have utilized the seasonal weather forecasts datasets from
 169 ECMWF System 5 (Johnson et al., 2019) in this study. Since 1993, ECMWF has been providing 51 ensemble
 170 forecasts (a set of multiple forecasts equally likely) on a monthly basis (25 ensembles prior to 2017) with a
 171 horizontal resolution of 1° × 1° and daily temporal resolution up to 7 months ahead. In this study, the time period
 172 from 1993 to 2020 was selected and the ensemble forecasts for the selected catchments have been downloaded
 173 from the CCDS. Here, we utilized data from 1993 to 2010 to generate bias correction factors, and data from 2011
 174 to 2020 to assess the skill (see Figure S1 in the supplementary material).

175 2.2 Methodology

176 The methodology of our analysis is summarized in the schematic diagram shown in Figure 2. Firstly, we compiled
 177 seasonal weather forecasts ensemble from ECMWF for precipitation (P), temperature (T), and PET over the 12
 178 reservoirs for 10 years from 2011 to 2020. To downscale the datasets, a linear scaling method was applied to each
 179 weather forcing (Sec. 2.2.1). Secondly, we estimated the parameters of the hydrological model and validated its

180 performance (Sec. 2.2.2). Utilizing the seasonal weather forecasts dataset as input data to the hydrological model,
 181 we generated an ensemble of SFFs, and using historical weather observations as input, we produced ESP.
 182 Specifically, to calculate ESP, 45 ensemble members of each weather variable were also selected from historical
 183 observations (1966-2010, see Figure S1). Each ensemble member represents the simulated flow using a
 184 hydrological model initialized with observed meteorological data to simulate current conditions and forced by
 185 historical meteorological observations for the forecasting period. The Continuous Ranked Probability Score
 186 (CRPS) and the Continuous Ranked Probability Skill Score (CRPSS) were applied (Sec. 2.2.3) to calculate the
 187 absolute performance (score) of each forecast product (Sec. 3.1 and 3.2) and the relative performance (overall
 188 skill) of SFFs with respect to ESP (Sec. 3.3, 3.4).



189
 190 **Figure 2: Schematic diagram illustrating analysis method of the study.**

191 Specifically, in Section 3.1, we analysed the contribution of hydrological modelling uncertainty to the
 192 performance of SFFs by comparing the actual score to the theoretical score, which is calculated using pseudo flow
 193 observations. Here, pseudo-observation refers to the flow time-series obtained by feeding the hydrological model
 194 with weather observations, i.e. where errors due to hydrological model structure are removed. In Section 3.2 we
 195 investigated which weather variable mostly influence the performance of SFFs. For doing so, we first calculated
 196 the ‘isolated score’ of the flow forecasts generated by forcing the hydrological model with seasonal weather
 197 forecasts for one meteorological variable while using observational data for the other two variables. For instance,
 198 to assess the contribution of precipitation, we calculated the isolated score-P using seasonal precipitation forecasts,
 199 and observations for temperature and PET. Then, we computed the ‘integrated score’ using seasonal weather
 200 forecasts for all three variables and determined the ‘relative scores’ for each variable as the ratio of the isolated
 201 score over the integrated score. This workflow is illustrated in Figure S2 (supplementary material). In Sections
 202 3.3 to 3.5, we examined the regional and seasonal variations and the characteristics of overall skill under extreme
 203 climate conditions.

204 **2.2.1 Bias correction (Statistical downscaling)**

205 The seasonal weather forecasts datasets from CCDS have a spatial resolution of $1^{\circ} \times 1^{\circ}$, which is too coarse for the
 206 catchment-scale analysis. Previous studies also have reported that seasonal weather forecasts generated from
 207 General Circulation Models contain systematic biases and this can cause forecast uncertainty (Manzanas et al.,
 208 2017; Maraun, 2016; Tian et al., 2018). Moreover, the usefulness of bias correction in enhancing the forecast skill
 209 has been shown in many previous studies (Crochemore et al., 2016; Ferreira et al., 2022, Pechlivanidis et al., 2020;
 210 Tian et al., 2018). Hence, it is imperative to investigate the potential enhancement in the skill of hydrological
 211 forecasts resulting from the bias correction of weather forcings.

212 Numerous bias correction methods have been developed including linear scaling method, local intensity scaling
 213 and quantile mapping (Fang et al., 2015; Shrestha et al., 2017). Thanks to its simplicity and low computation cost
 214 (Melesse et al., 2019), the linear scaling method is widely adopted. Despite its simplicity, this method has
 215 demonstrated practical usefulness in various studies (Azman et al., 2022; Crochemore et al., 2016; Shrestha et al.,
 216 2017), including our previous study on seasonal precipitation forecasts across South Korea (Lee et al., 2023).
 217 Therefore, the linear scaling method was utilized in this study.
 218 Previous studies found that additive correction is preferable for temperature whereas multiplicative correction is
 219 preferable for variables such as precipitation, evapotranspiration, and solar radiation (Shrestha et al., 2016).
 220 Consequently, the equations for linear scaling method for each variable can be expressed as:

$$221 \quad P_{forecasted}^* = P_{forecasted} \cdot (b_P)_m = P_{forecasted} \cdot \left[\frac{\mu_m(P_{observed})}{\mu_m(P_{forecasted})} \right] \quad (1)$$

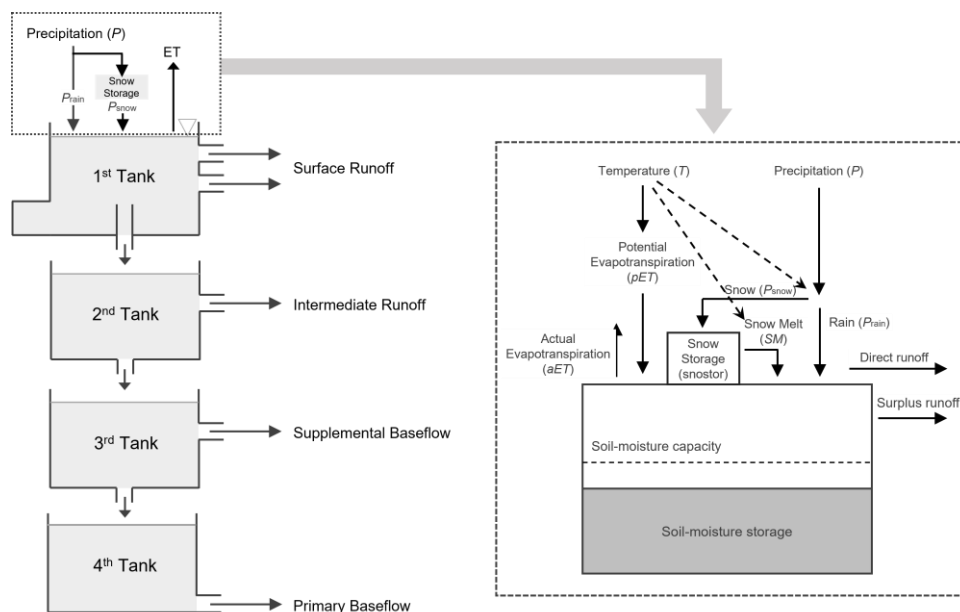
$$222 \quad PET_{forecasted}^* = PET_{forecasted} \cdot (b_{PET})_m = PET_{forecasted} \cdot \left[\frac{\mu_m(PET_{observed})}{\mu_m(PET_{forecasted})} \right] \quad (2)$$

$$223 \quad T_{forecasted}^* = T_{forecasted} + (b_T)_m = T_{forecasted} + \left[\mu_m(T_{observed}) - \mu_m(T_{forecasted}) \right] \quad (3)$$

224 where $X_{forecasted}^*$ is the bias corrected forecast variable (X) at daily time scale, $Y_{forecasted}$ is the original forecast
 225 variable (Y) before bias correction, $(b_Y)_m$ is the bias correction factors for each variable at month m . μ_m represents
 226 monthly mean, and $Y_{observed}$ is the observed daily data for the variable (Y). In this study, daily precipitation
 227 forecasts were bias corrected using the monthly bias correction factor (b_m) for each month ($m = 1$ to 12). The bias
 228 correction factor was computed using the observations and original forecasts datasets from 1993 to 2010, and
 229 these were then applied to adjust each seasonal weather forecasts for later years (2011 to 2020).

230 2.2.2 Hydrologic modelling

231 The Tank model was first developed by Sugawara of Japan in 1961 (Sugawara et al., 1986; Sugawara, 1995) and
 232 has become a widely used conceptual hydrologic model in many countries (Goodarzi et al., 2020; Ou et al. 2017).
 233 A modified version of the Tank model, incorporating soil moisture structure and snowmelt modules, is commonly
 234 used in South Korea for long-term water resources planning and management purposes due to its higher
 235 performance (Kang et al., 2004; Lee et al., 2020). As shown in Figure 3, the modified Tank model used in this
 236 study comprises four storage tanks representing the runoff and baseflow in the target catchment (Phuong et al.,
 237 2018; Shin et al., 2010) and incorporates a water-balance module suggested by the United States Geological
 238 Survey (McCabe and Markstrom, 2007).



239
 240 **Figure 3: The structure of modified Tank model (Left) and its water -balance module (Right)**

241 This model has 21 parameters (see Table S1 in the supplementary material), which were calibrated based on
 242 historical observations. We calibrated the model using observations for the period from 2001 to 2010, and the
 243 validation was done using the time period 2011 to 2020. To estimate the model parameters, the Shuffled Complex
 244 Evolution global optimization algorithm (SCE-UA), developed at the University of Arizona (Duan et al., 1992,
 245 1994), is utilized. This algorithm has widely been used for the calibration of hydrologic models and has shown
 246 more robust and efficient performance compared to many traditional optimization methods such as Genetic
 247 Algorithm, Differential Evolution, and Simulated Annealing (Rahnamay-Naeini et al., 2019; Yapo et al., 1996).
 248 The following Objective Function (OF) proposed by Sugawara (Sugawara et al., 1986), is applied for the SCE-
 249 UA algorithm, because a previous study demonstrated that this objective function shows superior results in
 250 calibrating the Tank model in South Korean catchments with calibration periods longer than 5 years (Kang et al.,
 251 2004).

$$252 \quad OF = \sum_{t=1}^N |q_t^{obs} - q_t^{sim}| / q_t^{obs} \quad (4)$$

253 where t , N represent time (in days) and total number of time steps, q_t^{obs} and q_t^{sim} represent the observed and
 254 simulated flow at time t , respectively. The optimal parameter set is the one that produces the lowest value from
 255 the objective function.

256 In order to evaluate the model performance in diverse perspectives, we used three different evaluation indicators:
 257 Nash-Sutcliffe model Efficiency coefficient (*NSE*), Percentage Bias (*PBIAS*), and Ratio of Volume (*ROV*). The
 258 calculation of each indicator was carried out as described by the following equations.

$$259 \quad NSE = 1 - \sum_{t=1}^N (q_t^{obs} - q_t^{sim})^2 / \sum_{t=1}^N (q_t^{obs} - q_{mean}^{obs})^2 \quad (5)$$

$$260 \quad PBIAS = \sum_{t=1}^N (q_t^{obs} - q_t^{sim})^2 / \sum_{t=1}^N q_t^{obs} \times 100 \quad (6)$$

$$261 \quad ROV = \sum_{t=1}^N q_t^{sim} / \sum_{t=1}^N q_t^{obs} \quad (7)$$

262 where t , N , q_t^{obs} and q_t^{sim} are as defined in Eq.4, and q_{mean}^{obs} represents observed mean flow across the total
 263 number of time steps (N).

264 The *NSE* can range from negative infinity to 1. A value of 1 indicates a perfect correspondence between the
 265 simulated and the observed flow. *NSE* values between zero and 1 are generally considered acceptable levels of
 266 performance (Moriasi et al., 2007). *PBIAS* is a metric used to measure the average deviation of the simulated
 267 values from the observation data. The optimal value of *PBIAS* is 0, and low-magnitude values indicate accurate
 268 simulation. Positive (negative) values of *PBIAS* indicate a tendency for overestimation (underestimation) in the
 269 hydrologic modelling (Gupta et al., 1999). *ROV* represents the ratio of total volume between the simulated and
 270 observed flow. An optimal *ROV* value is 1, and a value greater (less) than 1 suggests overestimation
 271 (underestimation) of total flow volume (Kang et al., 2004).

272 2.2.3 Score and skill assessment

273 As a score metric, we adopted the CRPS developed by Matheson and Winkler (1976) which measures the
 274 difference between the cumulative distribution function of the forecast ensemble and the observations. The CRPS
 275 has the advantage of being sensitive to the entire range of the forecast and being clearly interpretable, as it is equal
 276 to the Mean Absolute Error for a deterministic forecast (Hersbach, 2000). For these reasons, it is a widely used
 277 metric to assess the performance of ensemble forecasts (Leutbecher and Haiden, 2020). The CRPS can be
 278 calculated as:

$$279 \quad CRPS = \int [F(x) - H(x \geq y)]^2 dx \quad (8)$$

280 where $F(x)$ represents the cumulative distribution of SFFs ensemble, x and y are respectively the forecasted and
 281 observed flow, H is called the indicator function and is equal to 1 when $x \geq y$ and 0 when $x < y$. If SFFs were
 282 perfect, i.e., all the ensemble members exactly matched the observations, the CRPS would be equal to 0.
 283 Conversely, a higher CRPS indicates a lower the performance, as it implies that the forecast distribution is further
 284 from the observation. Note that the CRPS measures the absolute performance (score) of forecast without
 285 comparing it to a benchmark.

286 Along with the CRPS, we also employed the CRPSS, which presents the forecast performance in a relative manner
 287 by comparing it to a benchmark forecasting method. It is defined as the ratio of the forecast and benchmark score
 288 and is expressed as follows:

$$289 \quad \text{CRPSS} = 1 - \frac{\text{CRPS}^{\text{Sys}}}{\text{CRPS}^{\text{Ben}}} \quad (9)$$

290 where CRPS^{Sys} is the CRPS of the forecasting system (SFFs in our case) and CRPS^{Ben} is the CRPS of the
 291 benchmark. The values of CRPSS can range from $-\infty$ to 1. A CRPSS value between 0 to 1 indicates that the
 292 forecasting system has skill with respect to the benchmark. Conversely, when the CRPSS is negative, i.e., from $-\infty$
 293 to 0, the system gives lower performances than the benchmark. Here, we utilized ESP as a benchmark due to
 294 its extensive application in flow forecasting (Pappenberger et al., 2015; Peñuela et al., 2020) and its computational
 295 efficiency (Baker et al., 2021; Harrigan et al., 2018). ESP is generated using the Tank model fed with historical
 296 daily meteorological records from 1966 to 2010. As this period covers 45 years, ESP is composed of 45 members
 297 for each catchment.

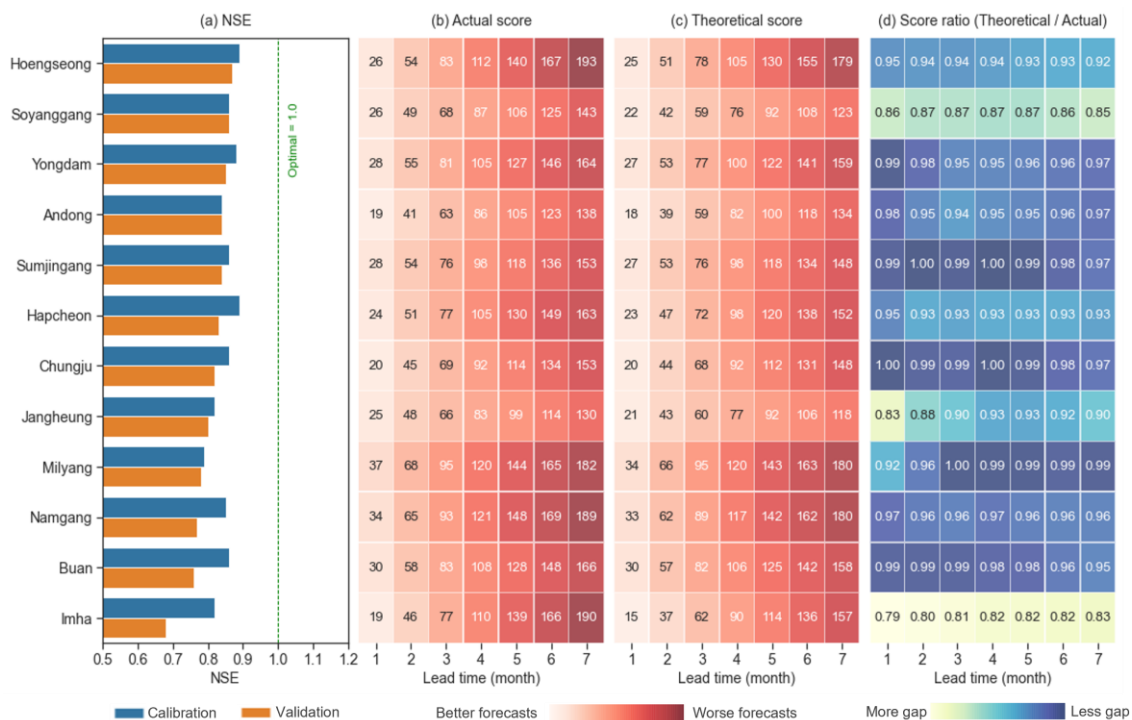
298 Since the CRPSS ranges from $-\infty$ to 1, simply averaging the CRPSS values over a period can result in low or no
 299 skills due to the presence few extremely negative values. To address this issue, here we employed the ‘overall
 300 skill’ metric introduced by Lee et al. (2023). The overall skill represents the probability with which a forecasting
 301 system (in our case, the SFFs) outperforms the benchmark (i.e., has CRPSS greater than 0) over a specific period.
 302 It is calculated as:

$$303 \quad \text{Overall skill (\%)} = \frac{\sum_{y=1}^{N_y} [H(\text{CRPSS}(y))] }{N_y} \times 100 (\%) \quad (10)$$

304 where N_y is the total number of years, the indicator function H is equal to 1 when $\text{CRPSS}(y) > 0$ (SFFs have skill
 305 with respect to ESP in year y) and 0 when $\text{CRPSS}(y) \leq 0$ (ESP outperforms SFFs). If the overall skill is greater
 306 than 50%, we can conclude that SFFs generally have skill over ESP across the period.

307 3. Results

308 3.1 Contribution of hydrological model to the performance of SFFs



309
 310 **Figure 4. (a) Nash-Sutcliffe Efficiency (NSE) of the hydrological models for the 12 catchments analysed in this study;**
 311 **(b) actual score and (c) theoretical score of SFFs, (d) score ratio (theoretical / actual) in terms of mean CRPS at different**

312 lead times (x-axis) (the scores are calculated before the bias correction of weather forcings). The actual score is
313 determined by comparing SFFs to flow observations. The theoretical score is determined by comparing SFFs to pseudo-
314 observations produced by the same hydrological model forced with observed precipitation, temperature and PET.

315 Figure 4(a) shows the NSE of the modified Tank model for each catchment during the calibration period 2001-
316 2010 (blue bars) and the validation period 2011-2020 (orange bars). As seen in this figure, the NSE values for the
317 12 catchments are generally high (within the range of 0.7 to 0.9) during both the calibration and validation periods,
318 and the relative difference in performance between the two periods is small for all catchments. Specifically, the
319 NSE results indicate a 'good' performance through comparative analysis (Chiew and McMahon, 1993; Moriasi et
320 al., 2015). However, the last three catchments (Namgang, Buan and Imha) exhibit a relatively greater gap between
321 calibration and validation periods. Among all 12 catchments, these three exhibit the most distinctive hydrological
322 characteristics: Imha is the driest, while Namgang is the wettest catchment, and Buan is located along the coast,
323 with the smallest catchment area. A detailed model performance evaluation, including other metrics such as
324 PBIAS and ROV (refer to Figure S3 in the supplementary material), also supports this result. Overall, Figure 4
325 demonstrates that the Tank model utilized in this study shows an excellent performance in simulating flow, with
326 relatively higher modelling challenges observed in those three catchments.

327 Figures 4(b-c) represent the actual and theoretical scores (mean CRPS) over the period 2011-2020. Again, these
328 are calculated by comparing the simulated flows with the observed flows (actual score), and with pseudo-
329 observations (theoretical score), respectively. Since the CRPS is computed based on accumulated monthly mean
330 flow at a given lead time, forecast errors also accumulate over time. Therefore, both scores deteriorate
331 considerably as the lead time increases. Generally, the theoretical scores are slightly smaller than the actual scores,
332 but the difference is marginal.

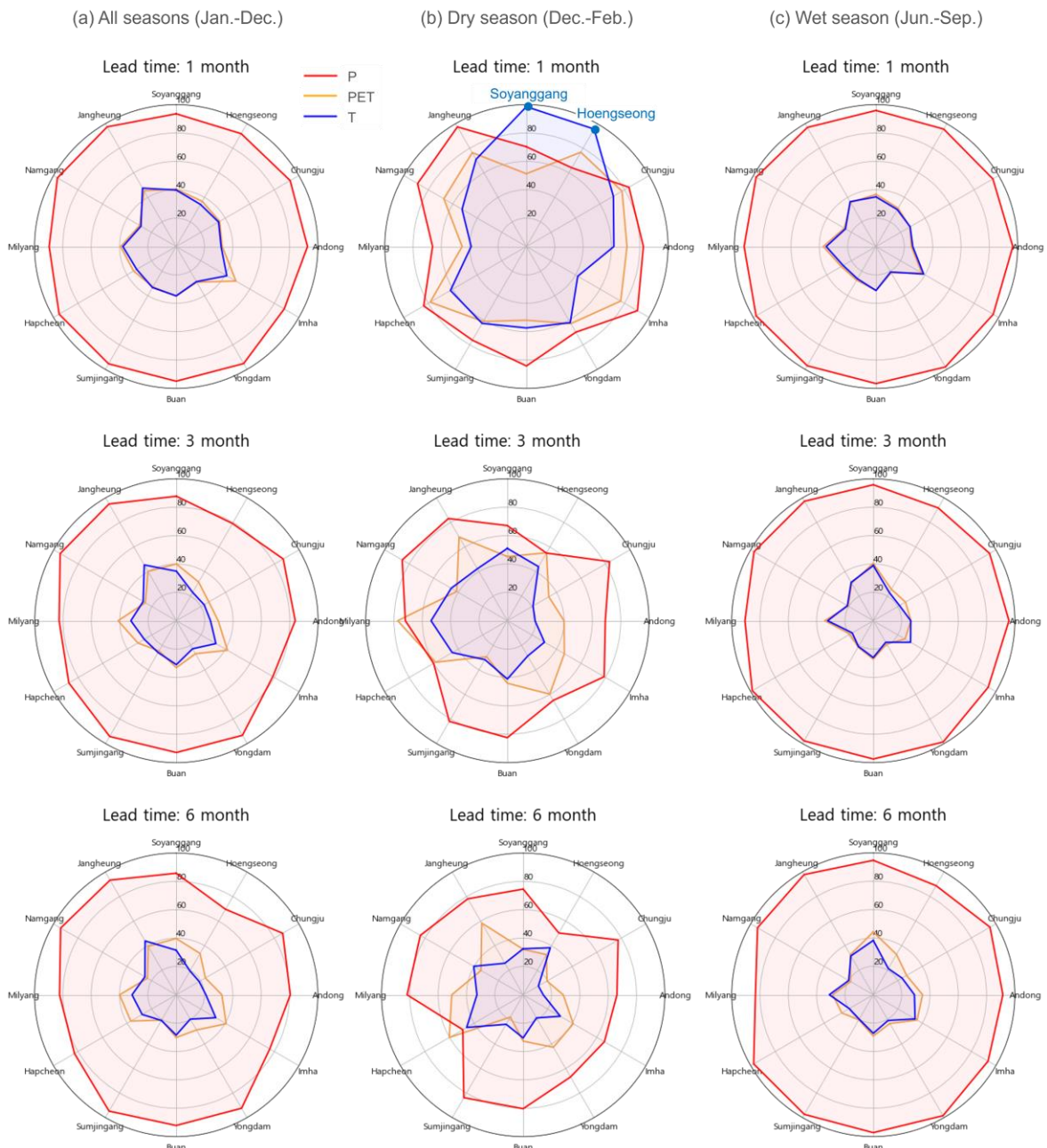
333 To facilitate comparison, the ratio between the actual score and theoretical score is shown in Figure 4(d). For most
334 catchments, the ratio values are close to 1, confirming the small gap between actual and theoretical score. The
335 noticeable exception is only seen in Imha catchment, characterised by being the driest among the catchments and
336 exhibiting the lowest modelling performance (Figure 4(a)).

337 3.2 Contribution of weather forecasting to the performance of SFFs

338 In this section, we quantify the contribution of each weather forcing forecast to the performance of SFFs, as
339 measured by the CRPS (see Section 2.2 and Figure S2 in the supplement material for details on the underpinning
340 methodology). Figure 5 shows the relative scores for each non-bias corrected weather forecast across all seasons
341 (a), dry season (b) and wet season (c) at different lead times (1, 3, and 6 months). The relative score is calculated
342 as the ratio of the integrated score (computed using seasonal weather forecasts for all weather forcings), to the
343 isolated score (when SFFs are computed using seasonal forecasts for one weather forcing, and observations for
344 the other two). The closer the isolated score to the integrated score, the larger the contribution of that weather
345 forcing to the overall performance (or lack of performance) of the SFFs.

346 As shown in Figure 5(a), the contribution of each weather forecast to the performance of SFFs varies with
347 catchment and lead time, but overall precipitation forecast plays a dominant role. Specifically, the contribution of
348 precipitation forecast (red) accounts for almost 90% of the integrated score which forced by seasonal weather
349 forecasts for all weather forcings, while PET (orange) and temperature (blue) contribute a similar level, ranging
350 between 30% and 40%.

351 During the dry season (Figure 5(b)) however, PET and temperature show comparable levels of contribution to
352 precipitation. This is more evident in the Soyanggang and Hoengseong catchments, which are both located in the
353 northernmost region of South Korea (see Figure 1). These catchments are characterized by low temperatures and
354 heavy snowfall in the dry (winter) season. Correct prediction of temperature is thus crucial here as temperature
355 controls the partitioning of precipitation into rain and snow, and hence the generation of a fast or delayed flow
356 response. Further analysis (shown in the supplementary material, Figure S4), reveals that temperature forecasts in
357 these two catchments are consistently lower than observation, which means that the hydrological model classifies
358 rain as snow for several events, and hence retains that 'snow' in the simulated snowpack which in reality should
359 produce a flow response. This explains the significant increase in performance when forcing the model with bias
360 corrected temperature instead (Figure S4(b)).



361

362
363
364
365

Figure 5. Relative score (%) of each weather forcings (Precipitation: red, PET: orange, Temperature: blue) before bias correction to the score of SFFs averaged over 10 years (2011-2020) during (a) all seasons, (b) dry and (c) wet season at 1, 3 and 6 lead months from the top to bottom (Catchments are ordered by their location from the northernmost (Soyanggang) to the southernmost (Jangheung) in right-angle direction, see Figure 1).

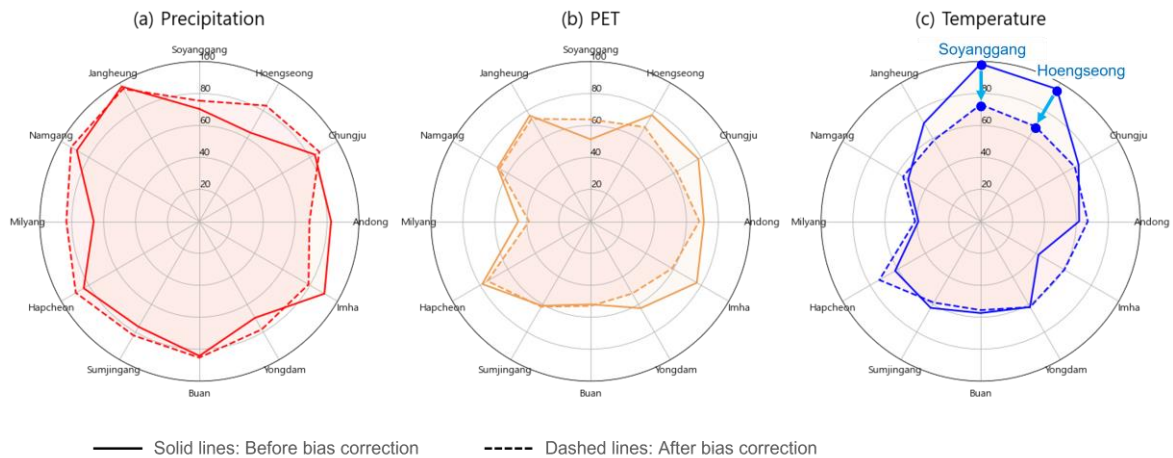
366
367
368
369
370
371

In order to enhance the forecasting performance, we applied bias correction to each weather forcing and re-generated SFFs with bias-corrected weather forcings. In most catchments and lead times, the overall skill is improved after correcting biases. The overall skill increases by 46% to 54% on average across all seasons, and more specifically from 31% to 50% in the dry season and from 54% to 55% in the wet season. The largest increase in overall skill is found in the Imha catchment, which had the lowest skill before correcting biases. For a detailed account of overall skill before and after bias correction, see Figure S5 and S6 in the supplementary material.

372
373
374
375
376

Figure 6 illustrates the change in the relative score of each weather forcing after bias correction, focusing on the dry season and the first forecasting lead month. One notable finding is that, in the snow-affected catchments (Soyanggang and Hoengseong), there is a significant decrease in the relative score of temperature after applying bias correction. As shown in detail in Figure S4 in the supplementary material, this is due to the correction of systematic underestimation biases in temperature forecasts, which lead to a more correct partition of precipitation

377 into snow and rain, and thus better flow predictions. The relative score of the forecasts for all seasons and lead
 378 times after bias correction is reported in Figure S7 in the supplementary material.

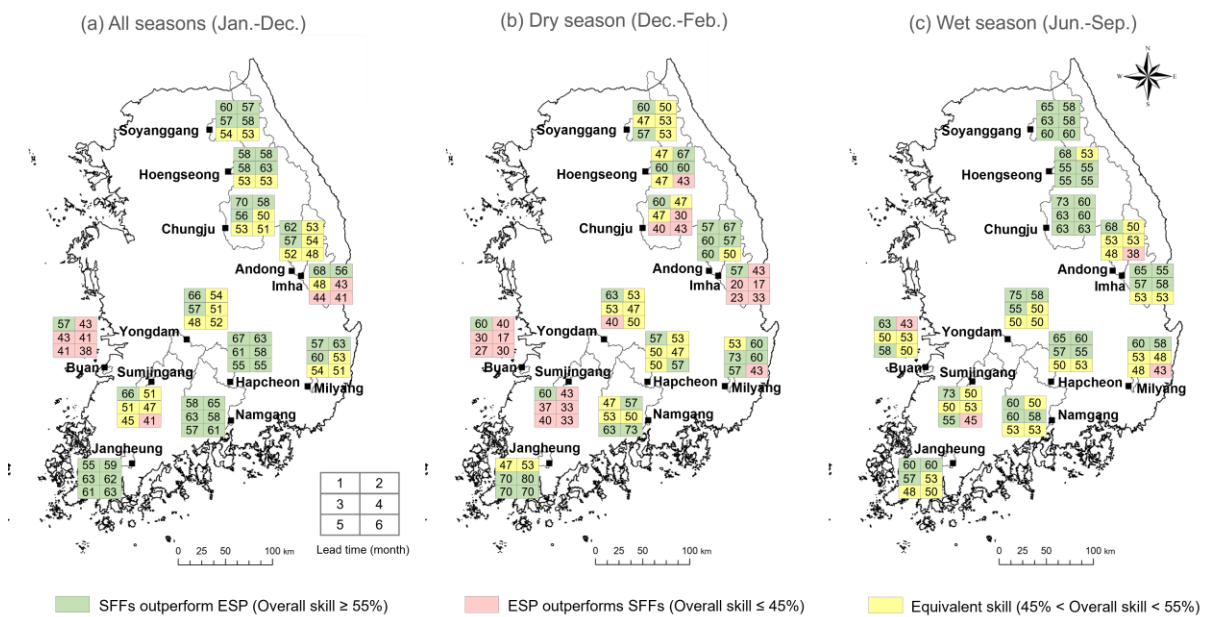


379 ——— Solid lines: Before bias correction - - - - - Dashed lines: After bias correction

380 **Figure 6. Relative score (%) of each weather forcings ((a) Precipitation, (b) PET, (c) Temperature), before (solid line)**
 381 **and after (dashed line) bias correction, to the score of SFFs averaged over 10 years (2011-2020) during the dry season**
 382 **and first lead month.**

383 **3.3 Comparison between SFFs and ESP across seasons and catchments**

384 In order to comprehensively compare the performance of SFFs and ESP, we employed the overall skill, which
 385 quantifies the frequency with which SFFs outperform ESP, as outlined in section 2.2.3 (Eq.10). Figure 7 shows
 386 the seasonal and regional variations of overall skill (after bias correction) for all seasons (a), for the dry season (b)
 387 and the wet season (c). For each catchment, the results are visualised through a table showing the overall skill at
 388 lead times of 1 to 6 months. The table cells are coloured in green (pink) when SFFs outperform ESP (ESP
 389 outperforms SFFs). Yellow colour indicates that the system and benchmark have equivalent performance. In
 390 principle, this happens when the overall skill is equal to 50%, however in order to avoid misinterpreting small
 391 differences in overall skill, we classified all cases as equivalent when it is between 45% and 55%. While the choice
 392 of the range ($\pm 5\%$) is subjective, we find it helpful to assist analysts in avoiding spurious precision in a simple
 393 and intuitive manner.

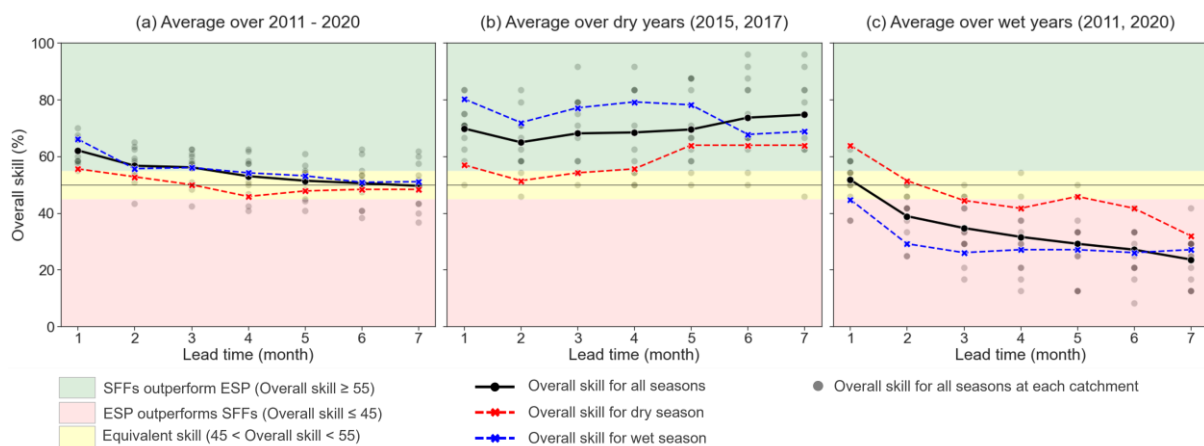


394 **Figure 7. Map of the overall skill of bias corrected SFFs for 10 years (2011-2020) over (a) all seasons, (b) dry season**
 395 **and (c) wet season. The colors represent whether SFFs outperform EPS or not for each catchment and lead time (1 to**
 396 **6 months).**

398 As shown in Figure 7(a), the overall skill of SFFs varies according to the lead time, season and catchment. SFFs
 399 generally outperform ESP, particularly up to 3 months ahead. At longer lead times, the results vary from catchment
 400 to catchment. For instance, in some catchments generally located in the Southern region, such as Janheung,
 401 Namgang, and Hapcheon, SFFs outperform ESP for longer lead times. On the other hand, in some catchments,
 402 such as Imha and Buan, ESP generally exhibits higher performance than SFFs. In specific, two catchments, Buan,
 403 which is located in the Western coastal region and has the smallest catchment area, and Imha, which is the driest
 404 catchment, show the lowest skill. Nevertheless, we could not identify a conclusive correlation between catchment
 405 characteristics such as size or mean annual precipitation and overall skill.
 406 Comparing the results for the dry and wet seasons, Figure 7(b-c) shows that SFFs are much more likely to
 407 outperform ESP in the wet season, and particularly in the catchments in northernmost region. During the dry
 408 season, overall skill of SFFs is lower, and particularly in the Buan, Imha and Sumjingang catchments SFFs
 409 outperform ESP only for the first lead month.

410 3.4 Comparison between SFFs and ESP in dry and wet years

411 We now assess the influence of exceptionally dry and wet conditions on the overall skill of SFFs. Based on the
 412 mean annual precipitation across 12 catchments within the period 2011-2020, we classified the years 2015 and
 413 2017 as dry ($P < 900$ mm), and the years 2011 and 2020 as wet ($P > 1500$ mm). Figure 8 shows the overall skill
 414 of SFFs averaged over 12 catchments for the entire period (a), dry years (b), and wet years (c), during all seasons
 415 (black solid line), dry (red dashed line) and wet (blue dashed line) seasons, respectively.



416
 417 **Figure 8. Overall skill of bias corrected SFFs over 12 catchments averaged over (a) all years (2011 to 2020), (b) dry**
 418 **years (mean annual $P < 900$ mm) and (c) wet years (mean annual $P > 1500$ mm) during all seasons (black lines), dry**
 419 **seasons (red dashed lines) and wet seasons (blue dashed lines). The pale black points represent the overall skill for all**
 420 **seasons at each catchment. Here, mean annual precipitation is averaged across the catchments and years.**

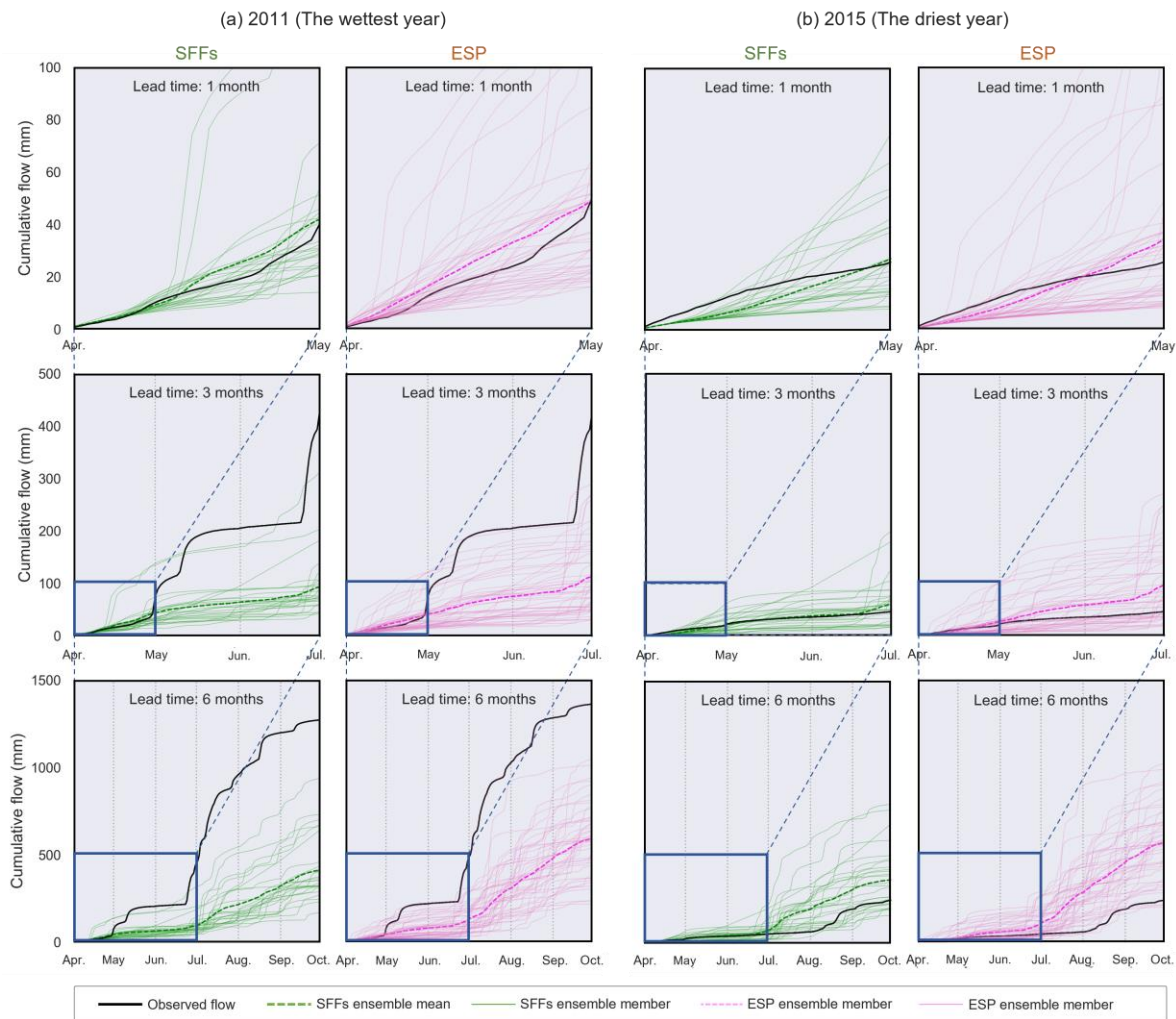
421 Figure 8(a) shows that SFFs outperform ESP for lead times of up to 3 months across seasons, while maintaining
 422 equivalent performance levels thereafter. In addition, it is evident that SFFs is more skilful during the wet season
 423 than during the dry season. In dry years (Figure 8(b)), in contrast to the typical decrease in the overall skill with
 424 lead time, we find that SFFs maintain a significantly higher skill at all lead times, and particularly during the wet
 425 season (blue line). On the other hand, in wet years (Figure 8(c)), the overall skill is generally poor, and ESP
 426 generally has higher performance than SFFs, especially during the wet season.

427 Last, we analyse the spatial variability of the overall skill by looking at the spread of individual catchments (grey
 428 dots). We see that the spread in dry and wet years (Figure 8(b-c)) is larger than in all years (Figure 8(a)). This
 429 confirms that under extreme weather conditions, the uncertainty and variability in the forecasting performance
 430 increase depending on the catchment. A more detailed analysis of the overall skill for each catchment (described
 431 in Figure S8 in the supplementary material) shows that the catchments located in Southern region consistently
 432 exhibit higher skill, regardless of lead times and dry/wet years.

433 3.5 Example of flow forecasts time-series

434 Figure 9 shows an example of the flow into the Chungju reservoir, which holds the largest storage capacity in
 435 South Korea. The overall skill of this catchment is the highest for a 1-month lead time; however, from the second
 436 lead month onward, it shows a moderate level of overall skill compared to other catchments (see Figure S8 in the
 437 supplementary material). In this section, we compare the observed and forecasted cumulative flow forced by

438 seasonal weather forecasts (SFFs, green lines) and historical weather records (ESP, pink lines) for lead times of
 439 1, 3, and 6 months from April during the wettest (2011) and the driest year (2015), respectively.



440
 441 **Figure 9. Observed cumulative flow (black lines) and forecasted cumulative flow representing SFFs after**
 442 **bias correction (left, green lines) and ESP (right, pink lines) in the Chungju reservoir for 1, 3, and 6 months**
 443 **of lead times over (a) the wettest year (2011, 1884mm/year), and (b) the driest year (2015, 742mm/year).**

444 In this specific catchment and years, SFFs show equivalent or slightly higher performance than ESP at a 1-month
 445 lead time. However, as the lead time increases, the performance of both methods tends to deteriorate. Essentially,
 446 it indicates an underestimation in the wettest year and an overestimation in the driest year. In particular,
 447 considerably higher performance was found in SFFs compared to ESP in the driest year (Figure 9(b)). On the
 448 other hand, it is obvious that both methods have insufficient sharpness in forecasting flow in the wettest years for
 449 lead times of 3 and 6 months.

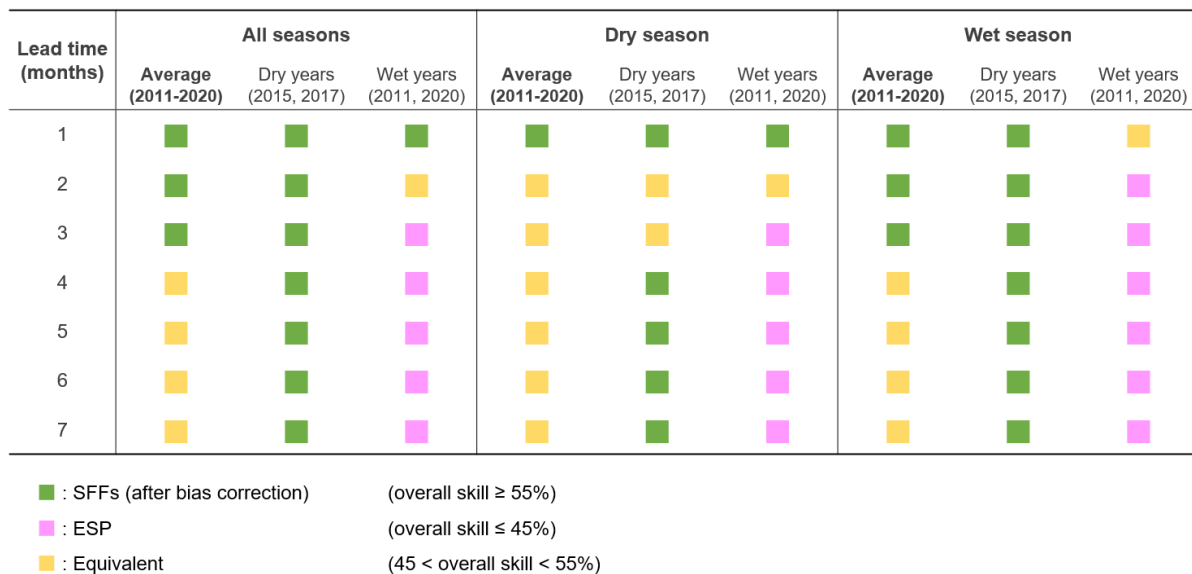
450 Examining each ensemble member of both SFFs and ESP, we found higher variability in ESP. Furthermore, since
 451 ESP utilizes the same weather forcings, the forecasted flows are generally similar in terms of its quantity and
 452 patterns, regardless of the wettest and driest years. Conversely, the forecasted flow ensemble members of SFFs
 453 show distinctive patterns for each year.

454 Although, these results are confined to a single catchment and specific years, this analysis is valuable in
 455 quantitatively illustrating the forecasted flow results under dry and wet conditions and different lead times.
 456 Furthermore, these features are generally shown in other catchments, and align with our previous findings in
 457 section 3.4.

458 4. Discussion

459 4.1 The skill of seasonal flow forecasts

460 This study offers a comprehensive view of overall skill of SFFs, benchmarked to the conventional – and easier to
 461 implement - ESP method. In contrast to the majority of previous studies, which assessed the skill of SFFs at
 462 continental or national level or over large river basins, our study focuses on 12 relatively small catchments (59 -
 463 6648 km²) across South Korea.



464

465 **Figure 10. Summary of key findings regarding the overall skill at different lead times, seasons, and years.**

466 Figure 10 summarizes the key findings of this study regarding the overall skill of SFFs across different seasons
 467 and years. It demonstrates that SFFs outperform ESP in almost all the cases for forecasting lead times of one
 468 month. This result is consistent with previous literature (e.g., Lucatero et al., 2018; Yossef et al., 2013). In addition,
 469 the higher skill of SFFs is also shown at lead times of 2 and 3 months in several situations as shown in Figure 10,
 470 and at even longer lead times in dry years. This is more surprising as this considerable performance of SFFs was
 471 not found in previous studies.

472 Similar to our study, earlier studies (Crochemore et al., 2016; Lucatero et al., 2018) have explored the skill
 473 compared with real flow observations at a catchment scale. Therefore, the comparison of their results with our
 474 findings holds interest. In brief, their results suggest that ESP remains a ‘hard-to-beat’ method compared to SFFs
 475 even after bias correction. Crochemore et al. (2016) showed that SFFs using bias corrected precipitation, is in
 476 equivalent level of performance with ESP up to 3 months ahead. Lucatero et al. (2018) concluded that SFFs still
 477 face difficulties in outperforming ESP, particularly at lead times longer than 1 month.

478 The difference of our results compared to the literature stems from a combination of several important factors.
 479 First, it is worth noting that these two previous studies were conducted at the catchment-scale, with a specific
 480 focus on Europe, namely France (Crochemore et al., 2016) and Denmark (Lucatero et al., 2018). The skill of SFFs
 481 varies according to the geographic locations, meteorological conditions of given study area, as confirmed by
 482 numerous studies (e.g., Greuell et al., 2018; Pechlivanidis et al., 2020; Yossef et al., 2013). Therefore, the skill of
 483 SFFs could also be influenced by distinct spatial and meteorological conditions between Europe and South Korea.
 484 Second, we can attribute the difference to the utilization of a more advanced seasonal weather forecasting system.
 485 Unlike previous studies which applied ECMWF system 4, our study is conducted based on ECMWF’s cutting-
 486 edge forecasting system version 5. It is reported that ECMWF system 5 has many improvements compared to the
 487 previous version including the predictive skill of the El Niño Southern Oscillation (ENSO) (Johnson et al., 2019)
 488 and rainfall inter-annual variability (Köhn-Reich and Bürger, 2019). Specifically, ENSO is known to be a key
 489 driver affecting the skill of seasonal weather forecasts (Ferreira et al., 2022; Shirvani and Landman, 2015;
 490 Weisheimer & Palmer, 2014); therefore, its improvement can result in notable changes in forecasting skill.
 491 Although the relationship between seasonal weather patterns in South Korea and ENSO is not fully understood,
 492 some previous research has shown good correlations for certain regions and seasons (Lee and Julien, 2016; Noh
 493 and Ahn, 2022). While it is challenging to quantitatively evaluate the impact of system advancements in this study,
 494 given the significance of meteorological forecast in hydrological forecasts, it is highly probable that the
 495 development of the system has had a positive influence on the results. Although a few studies have analysed the
 496 skill of SFFs based on ECMWF system 5 (e.g., Peñuela et al., 2020; Ratri et al., 2023), direct comparisons with

497 our research were deemed difficult due to differences in spatial scale and analysis methods, such as the absence
498 of a comparison with ESP.
499 Last, the performance of the hydrological model also contributes to differences in the results. To evaluate the
500 impact of hydrological model to SFFs, we compared actual score (forecast performance compared to observed
501 flow data) with theoretical score (forecast performance compared to pseudo flow observation) and found that the
502 actual scores are slightly higher than theoretical scores (i.e., theoretical score shows higher performance). This
503 finding is consistent with previous studies, and the gap between the actual and theoretical score is highly linked
504 to the performance of hydrological model (Greuell et al., 2018; van Dijk, 2013). When a model's actual score
505 closely approximates its theoretical score, it may suggest that the model is operating at a best possible level, given
506 the inherent uncertainties and limitations associated with the available data and methods. Although our results
507 demonstrated that the theoretical score shows higher performance than actual score, their difference was generally
508 marginal. This close agreement between the two scores indicates that the model is well-calibrated and capable of
509 effectively capturing the underlying hydrological processes in those catchments.

510 Our findings on the impact of bias correction quantitatively showed that generally precipitation controls the
511 performance of SFFs, however, we also found that temperature plays a substantial role in specific seasons and
512 catchments. Specifically, the Hoengseong and Soygang catchments, located in the northernmost part of South
513 Korea and affected by snowfall in the Dry (winter) season (December to February), exhibit a higher temperature
514 contribution than precipitation for a forecasting lead time of one month during the dry season. The main reason
515 for this is the underestimation of temperature forecasts. Our supplementary experiments provide evidence that
516 using bias-corrected temperature forecasts significantly improves the performance of flow forecasts (see Figure
517 S4 in the supplementary material). Although the positive impact of bias correction of precipitation forecasts in
518 enhancing the performance of SFFs has been well-documented in numerous previous studies (Crochemore et al.,
519 2016; Lucatero et al., 2018; Pechlivanidis et al., 2020; Tian et al., 2018), our result demonstrates the importance
520 of bias correction of temperature too, at least in snow-affected catchments.

521 An alternative approach to bias correction has been proposed by (Lucatero et al., 2018; Yuan and Wood, 2012),
522 who argue that directly correcting the biases in the flow forecasts may result in better performance at a lower
523 computational cost. However, we tested this approach and found conflicting outcomes (Figure S9 in the
524 supplementary material). Therefore, caution should be exercised when directly correcting biases for flow, as this
525 approach may exclude the contribution of initial conditions, which is one of the most crucial factors in
526 hydrological modelling. In cases where the performance of hydrological model is the major source of error, bias
527 correction of the flow might be useful; however, if the model shows an acceptable performance, as demonstrated
528 in this study, incorporating bias correction for the simulated flow could add more errors.

529 Due to limited data availability, conducting additional validation across a larger number of extreme events is not
530 possible. Nevertheless, our research findings suggest a potential correlation between the overall skill and dry/wet
531 conditions, that should be further validated if new data become available. Specifically, in the period analysed here,
532 SFFs considerably outperform ESP for all lead times during the wet season in dry years. Conversely, the overall
533 skill during the wet season in wet years was not satisfactory. This is because the overall skill is commonly
534 dominated by precipitation forecasting skill, and we previously found that the skill of precipitation forecasts is the
535 lowest in wet years (Lee et al 2023). The systematic biases of seasonal precipitation forecasts, which tend to
536 underestimate (overestimate) the precipitation during the wet (dry) season, led to the consistent results in flow
537 forecasts. This finding also hints that SFFs hold the potential to provide valuable information for effective water
538 resources management during dry conditions, which is crucial for drought management.

539 **4.2 Limitations and directions for future research**

540 In this paper, we investigated the overall skill of SFFs at the catchment scale using ECMWF's seasonal weather
541 forecasts (system 5) with a spatial resolution of $1 \times 1^\circ$. Based on our previous research, it has been demonstrated
542 that among four forecasting centres, ECMWF provides the most skilful seasonal precipitation forecasts (Lee et
543 al., 2023), thus we utilized seasonal weather forecasts datasets from ECMWF in this study. However, the skill for
544 other weather forcings such as temperature and PET, have not been tested across South Korea. Additionally, while
545 ECMWF provides seasonal weather forecasts with high resolution ($36 \times 36 \text{ km}$, approximately $0.3 \times 0.3^\circ$), we
546 utilized publicly available low resolution data ($1 \times 1^\circ$) to maintain consistency with our previous work (Lee et al.,
547 2023). Our additional investigation indicates that the difference in weather data between high and low resolution
548 is not substantial (see Figure S10 in the supplementary material). Nevertheless, prior studies suggest that the skill
549 of seasonal weather forecasts may vary according to factors such as region, season, and spatial resolution.
550 Therefore, broader research is required to determine the seasonal weather forecasts provider as well as spatial
551 resolution that can lead to skilful hydrological forecasts in the regions or seasons of interest.

552 Given the distinct climatic conditions in South Korea, it is important to acknowledge that our results may not be
553 applicable to other regions or countries. Therefore, further work needs to be carried out to reproduce this analysis
554 in different regions. To facilitate this process, two Python-based toolboxes can be useful: SEAFORM (SEAsonal
555 FORecasts Management) and SEAFLOW (SEAsonal FLOW forecasts). The SEAFORM toolbox, developed in
556 our previous study (Lee et al., 2023), offers multiple functions for manipulating seasonal weather forecast datasets
557 (e.g., download the datasets, time-series generation, bias correction). On the other hand, the SEAFLOW toolbox,
558 developed in this study, is specifically designed for the analysis of SFFs based on the modified Tank model (but
559 it could be useful to apply to other hydrologic models).
560 In terms of forecast skill, our study highlights the potential of SFFs at the catchment scale for real water resources
561 management. Nevertheless, it is crucial to recognize the difference between ‘skill’, indicating how well
562 hydrological forecasts mimic observed data, and ‘value’, referring to the practical benefits obtained from utilizing
563 those forecasts in real world. Previous studies have addressed this issue, showing that better skill does not always
564 result in higher value (Boucher et al., 2012; Chiew et al., 2003). While earlier findings suggest that the
565 conventional method (ESP) generally outperforms SFFs in terms of ‘skill’ (e.g., Lucatero et al., 2018; Yossef et
566 al., 2013), recent research demonstrates that, in terms of ‘value,’ the use of seasonal forecasts in semi-arid regions
567 offers significant economic benefits by mitigating hydro-energy losses in a dry year (Portele et al., 2021).
568 Therefore, our future research efforts should concentrate on a quantitative evaluation of the value of SFFs for
569 practical reservoir operations, informing decision-making in water resources management. This evaluation is of
570 significant importance as it directly relates to assessing the potential utilization of SFFs in practical water
571 management.

572 5. Conclusions

573 This study assessed the overall skill of SFFs across 12 catchments in South Korea using a hydrological model
574 forced by seasonal weather forecasts from the ECMWF (system 5). By focusing on operational reservoir
575 catchments with relatively small sizes, our findings showed the potential of SFFs for practical water resources
576 management.

577 The results first demonstrate that the performance of the hydrological model is crucial in flow forecasting with
578 the Tank model used in this study exhibiting reliable performance. Secondly, precipitation emerges as a dominant
579 factor influencing the performance of SFFs compared to other weather forcings, and this is more evident during
580 the wet season. However, temperature can also be highly important in specific seasons and catchments, and this
581 result highlights the significance of temperature bias correction as the flow simulation with the bias-corrected
582 temperature provides higher performance. Third, at catchment scale, which is more suitable for water resources
583 management, bias corrected SFFs have skill with respect to ESP up to 3 months ahead. Notably, the highest overall
584 skill during the wet season in dry years highlights the potential of SFFs to add value in drought management.
585 Lastly, while our research emphasizes the superior performance of SFFs at the catchment scale in South Korea, it
586 is important to note that outcomes may vary depending on factors such as the type of seasonal weather forecasts
587 system used, the study area, and the performance of the hydrological model.

588 As seasonal weather forecasting technologies continue to progress, it is also crucial to concurrently pursue their
589 application and validation in flow forecasting. We hope that our findings contribute to the ongoing validation
590 efforts of the skill of SFFs across various regions and, furthermore, serve as a catalyst for their practical application
591 in real-world water management. At the same time, our proposed workflow and the analysis package we have
592 developed using Python Jupyter Notebook, can offer valuable support to water managers in gaining practical
593 experience to utilize SFFs more effectively.

594 *Code and data availability.* The SEAFLOW (seasonal flow forecasts) and SEAFORM (seasonal forecast
595 management) Python packages are available at <https://github.com/uobwatergroup/seaflow>, and
596 <https://github.com/uobwatergroup/seaform>, respectively. ECMWF’s seasonal weather forecasts data are available
597 under a range of licences from <https://cds.climate.copernicus.eu/>. Reservoir and flow data are made available by
598 the K-water and can be downloaded from <https://www.water.or.kr/>.

599 *Author contributions.* YL designed the experiments, with suggestions from the other co-authors. YL developed
600 the workflow and performed simulation. FP and MAR participated in repeated discussions on interpretations of
601 results and suggested ways forward in the analysis. AP provided YL with modelling technical support and
602 reviewed the manuscript.

603 *Competing interests.* The authors declare that they have no conflict of interest.

604 *Acknowledgements.* Yongshin Lee is funded through a PhD scholarship by K-water (Korea Water Resources
605 Corporation). Andres Peñuela is funded by the European Research Executive Agency (REA) under the

606 HORIZON-MSCA-2021-PF-01 grant agreement 101062258. We also thank K-water and Dr. Shinuk Kang (South
607 Korea, K-water Institute) for sharing data and hydrological model (modified Tank) applied in this study.

608 References

- 609 Alley, R.B., Emanuel, K.A. and Zhang, F.: Advances in weather prediction. *Science*, [online] **363**, 342–344,
610 <https://doi.org/10.1126/science.aav7274>, 2019.
- 611 Arnal, L., Cloke, H.L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B. and
612 Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*,
613 **22**, 2057–2072, <https://doi.org/10.5194/hess-22-2057-2018>, 2018.
- 614 Azman, A.H., Tukimat, N.N.A. and Malek, M.A.: Analysis of Linear Scaling Method in Downscaling
615 Precipitation and Temperature, *Water Resources Management*, **36**, 171–179, <https://doi.org/10.1007/s11269-021-03020-0>, 2022.
- 617 Baker, S.A., Rajagopalan, B. and Wood, A.W.: Enhancing ensemble seasonal streamflow forecasts in the upper
618 Colorado river basin using multi-model climate forecasts, *Journal of the American Water Resources
619 Association*, **57**, 906–922, <https://doi.org/10.1111/1752-1688.12960>, 2021.
- 620 Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55,
621 <https://doi.org/10.1038/nature14956>, 2015.
- 622 Boucher, M.-A., Tremblay, D., Delorme, L., Perreault, L. and Ancil, F.: Hydro-economic assessment of
623 hydrological forecasting systems, *Journal of Hydrology*, **416-417**, 133–144,
624 <https://doi.org/10.1016/j.jhydrol.2011.11.042>, 2012.
- 625 Chiew, F.H.S., Zhou, S.L. and McMahon, T.A.: Use of seasonal streamflow forecasts in water resources
626 management, *Journal of Hydrology*, **270**, 135–144, [https://doi.org/10.1016/S0022-1694\(02\)00292-5](https://doi.org/10.1016/S0022-1694(02)00292-5), 2003.
- 627 Chiew, F. and McMahon, T.: Assessing the adequacy of catchment streamflow yield estimates, *Soil Research*,
628 **31**, 665, <https://doi.org/10.1071/sr9930665>, 1993.
- 629 Crochemore, L., Ramos, M.-H. and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill
630 of seasonal streamflow forecasts, *Hydrology and Earth System Sciences*, **20**, 3601–3618,
631 <https://doi.org/10.5194/hess-20-3601-2016>, 2016.
- 632 Day, G.N.: Extended streamflow forecasting using NWSRFS, *Journal of Water Resources Planning and
633 Management*, **111**, 157–170, [https://doi.org/10.1061/\(asce\)0733-9496\(1985\)111:2\(157\)](https://doi.org/10.1061/(asce)0733-9496(1985)111:2(157)), 1985.
- 634 Duan, Q., Sorooshian, S. and Gupta, V.K.: Effective and efficient global optimization for conceptual rainfall-
635 runoff models, *Water Resources Research*, **28**, <https://doi.org/10.1029/91wr02985>, 1992.
- 636 Duan, Q., Sorooshian, S. and Gupta, V.K.: Optimal use of the SCE-UA global optimization method for
637 calibrating watershed models, *Journal of Hydrology*, **158**, 265–284, [https://doi.org/10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4), 1994.
- 639 Fang, G.H., Yang, J., Chen, Y.N. and Zammit, C.: Comparing bias correction methods in downscaling
640 meteorological variables for a hydrologic impact study in an arid area in China. *Hydrology and Earth System
641 Sciences*, **19**, 2547–2559, <https://doi.org/10.5194/hess-19-2547-2015>, 2015.
- 642 Ferreira, G.W.S., Reboita, M.S. and Drumond, A.: Evaluation of ECMWF-SEAS5 seasonal temperature and
643 precipitation predictions over South America, *Climate*, **10**, 128, <https://doi.org/10.3390/cli10090128>, 2022.
- 644 Goodarzi, M., Jabbarian Amiri, B., Azarneyvand, H., Khazae, M. and Mahdianzadeh, N.: Assessing the
645 performance of a hydrological Tank model at various spatial scales, *Journal of Water Management Modeling*,
646 <https://doi.org/10.14796/jwmm.c472>, 2020.
- 647 Greuell, W., Franssen, W.H.P. and Hutjes, R.W.A.: Seasonal streamflow forecasts for Europe – Part 2: Sources
648 of skill. *Hydrology and Earth System Sciences*, **23**, 371–391, <https://doi.org/10.5194/hess-23-371-2019>, 2019.
- 649 Greuell, W., Franssen, W.H.P., Biemans, H. and Hutjes, R.W.A.: Seasonal streamflow forecasts for Europe –
650 Part I: Hindcast verification with pseudo- and real observations, *Hydrology and Earth System Sciences*, **22**,
651 3453–3472, <https://doi.org/10.5194/hess-22-3453-2018>, 2018.
- 652 Gupta, H.V., Sorooshian, S. and Yapo, P.O.: Status of automatic calibration for hydrologic models: Comparison
653 with multilevel expert calibration, *Journal of Hydrologic Engineering*, **4**, 135–143,
654 [https://doi.org/10.1061/\(asce\)1084-0699\(1999\)4:2\(135\)](https://doi.org/10.1061/(asce)1084-0699(1999)4:2(135)), 1999.

- 655 Harrigan, S., Prudhomme, C., Parry, S., Smith, K. and Tanguy, M.: Benchmarking ensemble streamflow
656 prediction skill in the UK, *Hydrology and Earth System Sciences*, **22**, 2023–2039, [https://doi.org/10.5194/hess-](https://doi.org/10.5194/hess-22-2023-2018)
657 [22-2023-2018](https://doi.org/10.5194/hess-22-2023-2018), 2018.
- 658 Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for ensemble prediction
659 systems, *Weather and Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-](https://doi.org/10.1175/1520-0434(2000)015%3C0559:dotcrp%3E2.0.co;2)
660 [0434\(2000\)015%3C0559:dotcrp%3E2.0.co;2](https://doi.org/10.1175/1520-0434(2000)015%3C0559:dotcrp%3E2.0.co;2), 2000.
- 661 Jackson-Blake, L., Francois Clayer, Haande, S., James Edward Sample and S. Jannicke Moe: Seasonal
662 forecasting of lake water quality and algal bloom risk using a continuous Gaussian Bayesian network,
663 *Hydrology and Earth System Sciences*, **26**, 3103–3124, <https://doi.org/10.5194/hess-26-3103-2022>, 2022.
- 664 Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S.,
665 Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H. and Monge-Sanz, B.M.:
666 SEAS5: the new ECMWF seasonal forecast system, *Geoscientific Model Development*, **12**, 1087–1117,
667 <https://doi.org/10.5194/gmd-12-1087-2019>, 2019.
- 668 Kang, S.U., Lee, D.R. and Lee, S.H.: A study on calibration of Tank model with soil moisture structure, *Journal*
669 *of Korea Water Resources Association*, **37**, 133–144, 2004.
- 670 Kolachian, R. and Saghafian, B.: Deterministic and probabilistic evaluation of raw and post processed sub-
671 seasonal to seasonal precipitation forecasts in different precipitation regimes, *Theoretical and Applied*
672 *Climatology*, **137**, 1479–1493, <https://doi.org/10.1007/s00704-018-2680-5>, 2019.
- 673 Köhn-Reich, L. and Bürger, G.: Dynamical prediction of Indian monsoon: Past and present skill, *International*
674 *Journal of Climatology*, **39**, 3574–3581, <https://doi.org/10.1002/joc.6039>, 2019.
- 675 K-water (Korea Water Resources Corporation): *My water*, [online] www.water.or.kr. Available at:
676 <http://www.water.or.kr> [Accessed 4 Oct. 2022], 2022.
- 677 Lee, J.H. and Julien, P.Y.: Teleconnections of the ENSO and South Korean precipitation patterns, *Journal of*
678 *Hydrology*, **534**, 237–250, <https://doi.org/10.1016/j.jhydrol.2016.01.011>, 2016.
- 679 Lee, J.W., Chegal, S.D. and Lee, S.O.: A review of Tank model and its applicability to various Korean
680 catchment conditions, *Water*, **12**, 3588, <https://doi.org/10.3390/w12123588>, 2020.
- 681 Lee, Y., Peñuela, A., Pianosi F. and Rico-Ramirez, M.A.: Catchment-scale skill assessment of seasonal
682 precipitation forecasts across South Korea, *International Journal of Climatology*, **43**, 5092–5111,
683 <https://doi.org/10.1002/joc.8134>, 2023.
- 684 Leutbecher, M. and Haiden, T.: Understanding changes of the continuous ranked probability score using a
685 homogeneous Gaussian approximation, *Quarterly Journal of the Royal Meteorological Society*, **147**, 425–442,
686 <https://doi.org/10.1002/qj.3926>, 2020.
- 687 Li, H., Luo, L., Wood, E.F. and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal
688 hydrologic forecasting, *Journal of Geophysical Research*, **114**, <https://doi.org/10.1029/2008jd010969>, 2009.
- 689 Lucatero, D., Madsen, H., Refsgaard, J.C., Kidmose, J. and Jensen, K.H.: Seasonal streamflow forecasts in the
690 Ahlergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical
691 consistency, *Hydrology and Earth System Sciences*, **22**, 3601–3617, <https://doi.org/10.5194/hess-22-3601-2018>,
692 2018.
- 693 Manzanas, R., Lucero, A., Weisheimer, A. and Gutiérrez, J.M.: Can bias correction and statistical downscaling
694 methods improve the skill of seasonal precipitation forecasts? *Climate Dynamics*, **50**, 1161–1176,
695 <https://doi.org/10.1007/s00382-017-3668-z>, 2017.
- 696 Maraun, D.: Bias correcting climate change simulations - a critical review, *Current Climate Change Reports*, **2**,
697 211–220, <https://doi.org/10.1007/s40641-016-0050-x>, 2016.
- 698 Matheson, J.E. and Winkler, R.L.: Scoring rules for continuous probability distributions, *Management Science*,
699 **22**, 1087–1096, <https://doi.org/10.1287/mnsc.22.10.1087>, 1976.
- 700 McCabe, G.J. and Markstrom, S.L.: A monthly water-balance model driven by a graphical user interface, *U.S.*
701 *Geological Survey*, 1–2, 2007.
- 702 Melesse, A.M., Wossenu Abteu and Senay, G.: *Extreme hydrology and climate variability : monitoring,*
703 *modelling, adaptation and mitigation*, Amsterdam, Netherlands: Elsevier, 2019.

- 704 Ministry of Environment.: *2020 Korea annual hydrological report*. South Korea, Available at:
705 <https://www.mois.go.kr/firt/bbs/type001> (Accessed: 28 August 2022), 2020.
- 706 Ministry of land, infrastructure and transportation.: *Long-term Water Resources Management Master Plan*
707 (2001~2020), South Korea, 2016.
- 708 Moriasi, D. N., Gitau, M. W., Pai, N. and Daggupati, P.: Hydrologic and water quality models: Performance
709 measures and evaluation criteria, *Journal of the ASABE*, [online] **58**, 1763–1785,
710 <https://doi.org/10.13031/trans.58.10715>, 2015.
- 711 Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel R. D. and Veith T. L.: Model evaluation
712 guidelines for systematic quantification of accuracy in watershed simulations, *Journal of the ASABE*, **50**, 885–
713 900, <https://doi.org/10.13031/2013.23153>, 2007.
- 714 Noh, G.-H. and Ahn, K.-H.: Long-lead predictions of early winter precipitation over South Korea using a SST
715 anomaly pattern in the North Atlantic Ocean, *Climate Dynamics*, **58**, 3455–3469, [https://doi.org/10.1007/s00382-](https://doi.org/10.1007/s00382-021-06109-9)
716 021-06109-9, 2022.
- 717 Ou, X., Gharabaghi, B., McBean, E. and Doherty, C.: Investigation of the Tank model for urban storm water
718 management, *Journal of Water Management Modeling*, <https://doi.org/10.14796/jwmm.c421>, 2017.
- 719 Pappenberger, F., Ramos, M.H., Cloke, H.L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A. and Salamon,
720 P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, *Journal*
721 *of Hydrology*, **522**, 697–713, <https://doi.org/10.1016/j.jhydrol.2015.01.024>, 2015.
- 722 Pechlivanidis, I.G., Crochemore, L., Rosberg, J. and Bosshard, T.: What are the key drivers controlling the
723 quality of seasonal streamflow forecasts? *Water Resources Research*, **56**, 1–19,
724 <https://doi.org/10.1029/2019wr026987>, 2020.
- 725 Peñuela, A., Hutton, C. and Pianosi, F.: Assessing the value of seasonal hydrological forecasts for improving
726 water resource management: insights from a pilot application in the UK, *Hydrology and Earth System Sciences*,
727 **24**, 6059–6073, <https://doi.org/10.5194/hess-24-6059-2020>, 2020.
- 728 Phuong, H.T., Tien, N.X., Chikamori, H. and Okubo, K.: A hydrological Tank model assessing historical runoff
729 variation in the Hieu river basin, *Asian Journal of Water, Environment and Pollution*, **15**, 75–86,
730 <https://doi.org/10.3233/ajw-180008>, 2018.
- 731 Portele, T., Lorenz, C., Berhon Dibrani, Laux, P., Bliefernicht, J. and Kunstmann, H.: Seasonal forecasts offer
732 economic benefit for hydrological decision making in semi-arid regions, *Scientific Reports*, **11**, 10581,
733 <https://doi.org/10.1038/s41598-021-89564-y>, 2021.
- 734 Prudhomme, C., Hannaford, J., Alfieri, L., Boorman, D.B., Knight, J., Bell, V., Christopher A.-L. Jackson,
735 Svensson, C., Parry, S., Nuria Bachiller-Jareno, Davies, H., Davis, R.A., Mackay, J.D., Andrew, Rudd, A.C.,
736 Smith, K., Bloomfield, J.P., Ward, R. and Jenkins, A.: Hydrological outlook UK: an operational streamflow and
737 groundwater level forecasting system at monthly to seasonal time scales, *Hydrological Science Journal*, **62**,
738 2753–2768, <https://doi.org/10.1080/02626667.2017.1395032>, 2017.
- 739 Rahnamay-Naeini, M. Analui, B., Gupta, H.V., Duan, Q. and Sorooshian, S.: Three decades of the Shuffled
740 Complex Evolution (SCE-UA) optimization algorithm: Review and applications, *Scientia Iranica*, **26**, 2015–
741 2031, 2019.
- 742 Ratri, D.N., Weerts, A., Muharsyah, R., Whan, K., Tank, A.K., Aldrian, E. and Hariadi, M.H.: Calibration of
743 ECMWF SEAS5 based streamflow forecast in seasonal hydrological forecasting for Citarum river basin, West
744 Java, Indonesia, *Journal of Hydrology*, **45**, <https://doi.org/10.1016/j.ejrh.2022.101305>, 2023.
- 745 Shin, S.H., Jung, I.W. and Bae, D.H.: Study on estimation of optimal parameters for Tank model by using SCE-
746 UA, *Journal of Korea Water Resources Association*, 1530–1535, 2010.
- 747 Shirvani, A. and Landman, W.A.: Seasonal precipitation forecast skill over Iran, *International Journal of*
748 *Climatology*, **36**, 1887–1900. <https://doi.org/10.1002/joc.4467>, 2015.
- 749 Shrestha, M., Acharya, S.C. and Shrestha, P.K.: Bias correction of climate models for hydrological
750 modelling - are simple methods still useful? *Meteorological Applications*, **24**, 531–539,
751 <https://doi.org/10.1002/met.1655>, 2017.
- 752 Shrestha, S. and Htut, A.Y.: Modelling the potential impacts of climate change on hydrology of the Bago River
753 basin, Myanmar, *International Journal of River Basin Management*, **14**, 287–297,
754 <https://doi.org/10.1080/15715124.2016.1164177>, 2016.

- 755 Soares, M. B. and Dessai, S.: Barriers and enablers to the use of seasonal climate forecasts amongst
756 organisations in Europe, *Climatic Change*, **137**, 89–103, <https://doi.org/10.1007/s10584-016-1671-8>, 2016.
- 757 Sugawara, M., Watanabe, I., Ozaki, E. and Katsuyama, Y.: *Tank model programs for personal computer and the*
758 *way to use*, National Research Centre for Disaster Prevention, Japan, 1986.
- 759 Sugawara, M.: “*Tank model.*” *Computer models of watershed hydrology*, V. P. Singh (Ed.), Water Resources
760 Publications, Highlands Ranch, Colorado, 1995.
- 761 Tian, F., Li, Y., Zhao, T., Hu, H., Pappenberger, F., Jiang, Y. and Lu, H.: Evaluation of the ECMWF system 4
762 climate forecasts for streamflow forecasting in the upper Hanjiang river basin, *Hydrology Research*, **49**, 1864–
763 1879, <https://doi.org/10.2166/nh.2018.176>, 2018.
- 764 Van Dijk, A.I.J.M., Peña-Arancibia, J.L., Wood, E.F., Sheffield, J. and Beck, H.E.: Global analysis of seasonal
765 streamflow predictability using an ensemble prediction system and observations from 6192 small catchments
766 worldwide, *Water Resources Research*, **49**, 2729–2746, <https://doi.org/10.1002/wrcr.20251>, 2013.
- 767 Whateley, S., Palmer, R.N. and Brown, C.: Seasonal hydroclimatic forecasts as innovations and the challenges
768 of adoption by water managers, *Journal of Water Resources Planning and Management*, **141**,
769 [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000466](https://doi.org/10.1061/(asce)wr.1943-5452.0000466), 2015.
- 770 Weisheimer, A. and Palmer, T.N.: On the reliability of seasonal climate forecasts, *Journal of The Royal Society*
771 *Interface*, **11**, 20131162, <https://doi.org/10.1098/rsif.2013.1162>, 2014.
- 772 Yapo, P.O., Gupta, H.V. and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models:
773 sensitivity to calibration data, *Journal of Hydrology*, **181**, 23–48, [https://doi.org/10.1016/0022-1694\(95\)02918-](https://doi.org/10.1016/0022-1694(95)02918-4)
774 4, 1996.
- 775 Yoe, C.E.: *Principles of risk analysis : decision making under uncertainty*, Boca Raton, FL: Crc Press, Taylor
776 And Francis, 2016.
- 777 Yossef, N.C., Winsemius, H., Weerts, A., van Beek, R. and Bierkens, M.F.P.: Skill of a global seasonal
778 streamflow forecasting system, relative roles of initial conditions and meteorological forcing, *Water Resources*
779 *Research*, **49**, 4687–4699, <https://doi.org/10.1002/wrcr.20350>, 2013.
- 780 Yuan, X. and Wood, E.F.: Downscaling precipitation or bias-correcting streamflow? Some implications for
781 coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast, *Water Resources*
782 *Research*, **48**, 1–7, <https://doi.org/10.1029/2012WR012256>, 2012.