Exploring the value of seasonal flow forecasts for drought management in South Korea

Yongshin Lee[1], Andres Peñuela[2], Francesca Pianosi[1], Miguel Angel Rico-Ramirez[1]

[1] School of Civil, Aerospace and Design Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom

[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.

Drought poses significant challenges across various water-dependent sectors. In the past few decades, numerous devastating droughts have been reported worldwide including in South Korea. A recent drought in South Korea, which lasted from 2014 to 2016, led to significant consequences including water restrictions and nationwide crop failures. Historically, reservoirs have played a crucial role in mitigating hydrological droughts by increasing water supply stability. With exacerbating intensity and frequency of droughts, enhancing the operational efficiency of existing reservoirs becomes increasingly important. This study examines the value of Seasonal Flow Forecasts (SFFs) in informing reservoir operations during three historical drought events, with a focus on two key reservoir systems in South Korea. For these events, we simulate what would have happened if the reservoir managers had optimised operations using SFFs. For comparison, we also simulate the effect of reservoir operations optimised using two deterministic scenarios (worst-case and 20-year return period drought) and another ensemble forecasts product (Ensemble Streamflow Prediction, ESP). We repeat our simulation experiments by varying the key choices in the experimental set-up, i.e. the forecast lead time, decision-making time step, and the method for selecting a compromise solution between conflicting objectives. We then propose a new, simple and intuitive method for measuring the value of the different scenarios/forecasts, based on the frequency of outperforming (in a Pareto-dominance sense) the historical operation across such experiments. Our findings indicate that while deterministic scenarios show higher accuracy, forecast-informed operations with ensemble forecasts tend to yield greater value. This highlights the importance of considering forecast uncertainty in optimising reservoir operations. Although SFFs generally show higher accuracy than ESP, the difference in value is small. Last, sensitivity analysis shows that the method used to select a compromise release schedule between competing operational objectives is a key control of forecast value, implying that the benefits of using seasonal forecasts may vary widely depending on how priorities between objectives are established.

Keywords: drought, reservoir operations, seasonal weather forecasts, seasonal flow forecasts, ensemble streamflow prediction, multi-objective optimisation, multi-criteria decision-making

1. Introduction

Drought stands as one of the major natural disasters with devastating impacts for various sectors including agriculture, water resources, environment and energy (Mishra and Singh, 2010; Schwalm et al., 2017; Zhang et al., 2022). The severity of droughts is anticipated to escalate in the future under a warmer climate, but there is plenty of evidence to suggest that this increase may already be underway (Sheffield et al., 2012). In South Korea, a severe drought event, prolonged from 2014 to 2016, caused substantial consequences, such as water restrictions in certain regions and nationwide crop failures (K-water, 2018). Reservoirs have played a crucial role in mitigating drought impacts by stabilising water supply and compensating for hydrological variability (Goldsmith and Hildyard, 1984). However, the increasing frequency and intensity of extreme droughts are posing greater challenges for reservoir operators (Sheffield et al., 2012; Schwalm et al., 2017). On the other hand, the construction of new reservoirs has become increasingly controversial in many countries, including South Korea, mainly due to concerns about the socio-economic costs and undesirable environmental impacts of reservoir development (Ehsani et al., 2017). This highlights the growing significance of enhancing the operation of existing reservoirs to mitigate drought damages. A key contribution to this end may come by improving flow forecasting systems and their use in support of decision-making under extreme weather conditions (Turner et al., 2017).

Advancements in numerical weather prediction systems over the past decade have significantly improved forecasting performance at longer time scales (Bauer et al., 2015; Alley et al., 2019). Seasonal weather forecasts, which provide predictions of weather variables (e.g. precipitation, temperature) for the next several months, have gained interest among researchers for their potential in enhancing water resources management. Accordingly, numerous studies have been conducted to transform seasonal weather forecasts into more relevant Seasonal Flow Forecasts (SFFs) across various regions of the world (e.g. Prudhomme et al., 2017; Arnal et al., 2018; Greuell et
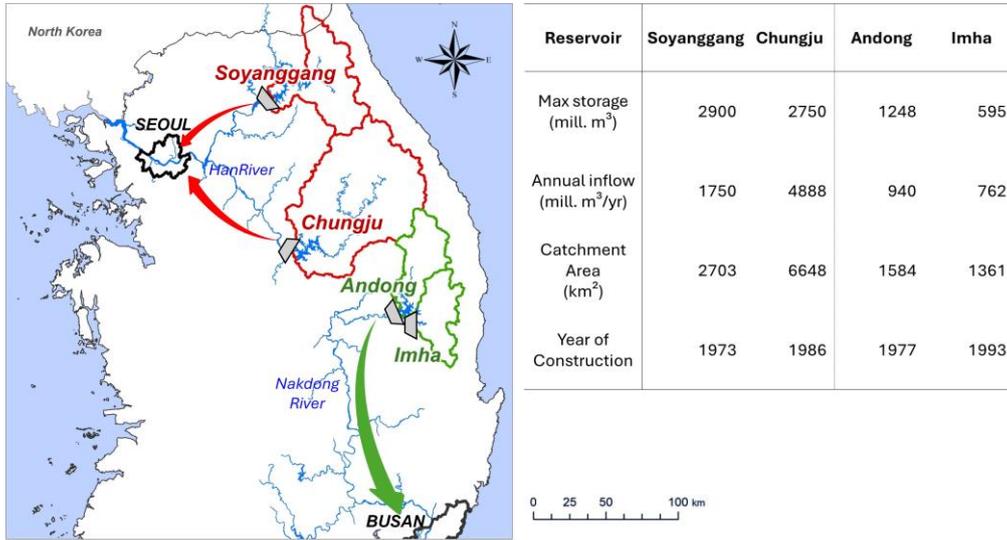
al., 2018; Lucatero et al., 2018; Hurkmans et al., 2023). In many countries, however, practical reservoir operations still make limited use of SFFs. Even when water resource modelling is used to inform operational decisions, reservoir operators tend to run these models against deterministic scenarios such as the worst-case scenario (Yoe, 2019) or against Ensemble Streamflow Prediction (ESP) (Day, 1985). The worst-case scenario mimics the most extreme historical event to hedge risks associated with uncertainties in water management, whereas ESP generates an ensemble of flow forecasts by forcing a hydrological model with historical meteorological observations (Baker et al., 2021). Previous studies have identified as primary obstacles to the use of SFFs by water managers their tendency to adopt a risk-adverse approach (Block, 2011), the lack of experience in handling SFFs products and the perceived low reliability of SFFs (Millner and Washington, 2011; Soares and Dessai, 2016; Jackson-Blake et al., 2022). Indeed, previous studies have shown that SFFs provide more accurate forecast than ESP only for the first or second months ahead (Yossef et al., 2013; Crochemore et al., 2016; Lucatero et al., 2018) and their performance decreases with increasing lead time (Greuell et al., 2018; Pechlivanidis et al., 2020).

In perspective of reservoir operations, however, more than the forecast accuracy, i.e. how well hydrological forecasts replicate observations, the attention should be directed to the forecast value, i.e. the benefits of using forecasts to inform operational decisions (Turner et al., 2017; Peñuela et al., 2020a). Assessing forecast value may reveal situations where using forecasts enhances water management even if the accuracy is relatively low (Rougé et al., 2023). With this idea in mind, several studies have utilised model simulations to assess how using SFFs could have improved reservoir operations during past events. To achieve this, these studies feed SFFs into a reservoir operation optimisation model to find the "optimal" release schedule based on those flow predictions and then assess the effects of the optimal schedule by simulating it against the actual observed flows. The process is repeated for as many decision time-steps as needed throughout the historical event. The performances of such forecast-informed operations are then summarised through a set of performance indicators and compared to the performances obtained with a benchmark approach, such as optimising against a deterministic scenario or using prescribed operation rules. The increase in performance with respect to the benchmark is regarded as the value of the SFFs. For example, Chiew et al. (2003) investigated the value of SFFs for agricultural supply from a reservoir in Australia, a region affected by El Nino/Southern Oscillation (ENSO) teleconnections. They found that release schedule informed by SFFs can yield modest benefits compared to using a predefined reservoir operation rule. Peñuela et al. (2020a) assessed the forecast value for reservoir operations in the UK, focusing on maximising water supply and minimising pumping energy cost. They found that using ensemble forecasts can significantly enhance operational efficiency compared to relying on a deterministic worst-case scenario, whereas ESP is a hard-to-beat benchmark. Crippa et al. (2023) assessed the value of SFFs for agricultural water supply in Greece and found that SFFs can marginally improve reservoir operations with respect to using a simple reservoir operation rule. However, they solely utilised the median of the SFFs ensemble, hence leaving open the question of whether using the full ensemble and allowing for uncertainty in the optimisation process could bring more obvious advantages, as found in Peñuela et al. (2020a).

This paper investigates the value of SFFs for informing reservoir operations in South Korea by assessing their potential to mitigate the impacts of three major historical drought events. We build on our previous works on the skill of seasonal precipitation and flow forecasts across catchments in South Korea, indicating that SFFs can be particularly suitable for predicting droughts. Specifically, in Lee et al. (2023), we compared the performance of precipitation forecasts from various forecasting centres and found that the European Centre for Mid-range Weather Forecasting (ECMWF) provides the most accurate forecasts in South Korea and particularly during dry years. Our subsequent research on translating seasonal weather forecasts into flow forecasts (Lee et al., 2024), demonstrated that SFFs are generally more accurate than ESP up to 3 months ahead and at even longer lead times in dry years. In this study, we focus on two reservoir systems, Soyanggang-Chungju and Andong-Imha, which serve as crucial water sources for the country, including densely populated metropolitan areas such as the capital city of Seoul, and three recent major droughts in 2001-02, 2008-09, and 2004-16. To identify the optimal 'forecast-informed' reservoir operations during these drought events, we employ a multi-objective optimisation approach driven by SFFs. For comparison, we also optimise against ESP and two deterministic scenarios currently utilised by the national water agency in charge of reservoir operations (K-water). To increase the robustness of our conclusions, simulation experiments are repeated with different choices of the forecast lead time, the method for selecting a compromise solution between the two conflicting objectives pursued by the reservoir managers (minimising short-term supply deficit versus maximising the storage volume at the end of the hydrological year), and the temporal resolutions for repeating the multi-objective optimisation, i.e. the decision-making time step. Finally, for each flow scenario/forecast, we synthetically measure the value of the scenario/forecast as the chances of achieving better operational outcomes compared to historical operations (i.e. Pareto-dominating historical operations) across all the simulation experiments. This new approach to measuring value is useful because it acknowledges the uncertainty in the simulation results due to experimental set-up choices while also capturing the trade-offs between the conflicting objectives in a simple, synthetic way.
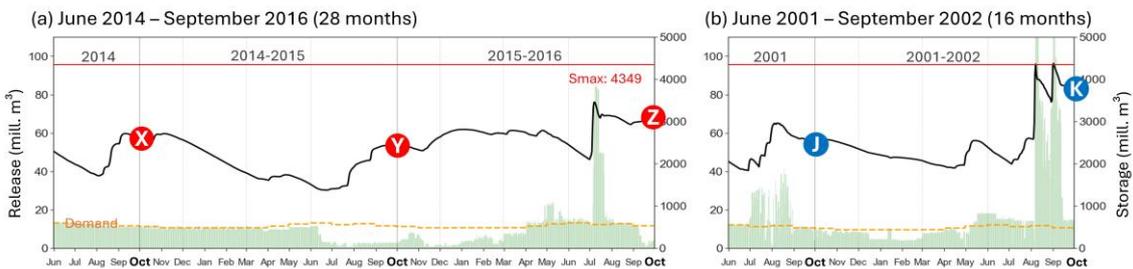
2. Study area and available data

2.1 Case study reservoirs and drought events

Figure 1: Location and properties of the studied reservoirs and their catchments. The green and red arrows represent the regions supplied by those reservoirs.

Reservoir	Soyanggang	Chungju	Andong	Imha
Max storage (mill. m³)	2900	2750	1248	595
Annual inflow (mill. m³/yr)	1750	4888	940	762
Catchment Area (km²)	2703	6648	1584	1361
Year of Construction	1973	1986	1977	1993

Figure 2: Daily reservoir operation records for the studied drought events (K-water, 2023). Points X, Y, Z, J, K, L, M represent the ends of the hydrological years (September 30th) which will be used as points in time for the forecast value assessment.

Currently, there are 20 multipurpose reservoirs in operation across South Korea, each playing a vital role in national water resources management and the mitigation of water-related disasters (Park and Kim, 2014). This study specifically focuses on two reservoir systems: Soyanggang-Chungju and Andong-Imha. The Soyanggang and Chungju reservoirs have the largest storage capacity in South Korea (Figure 1). They are positioned upstream of the Han River and serve as primary water sources for Seoul's metropolitan area, with a population of approximately 23 million people (K-water, 2023). In terms of total storage capacity, these two reservoirs also stand as the two largest across the country. The Andong and Imha reservoirs are located in northernmost region of the Nakdong River and supply water to plenty of cities alongside the river, including Busan, the second-largest city in the country.

3

132 The Soyanggang and Chungju reservoirs are operated conjunctively by the national water resources corporation,
133 K-water, effectively functioning as a single reservoir. For instance, during periods when one reservoir (e.g.
134 Soyanggang) experiences reduced storage volume, the other reservoir (e.g. Chungju) supplements the water
135 supply. A historical example of this conjunctive operations is provided in Figure S1 in the supplementary material.
136 A similar approach is used for the Andong and Imha reservoirs. Here again conjunctive operations are important
137 for mitigating drought damage. Therefore, in this study, we treat Soyanggang-Chungju as one integrated two-
138 reservoir system, and the same for Andong-Imha. The catchments feeding these systems exhibit similar
139 hydrological regimes, with approximately 70% of annual inflow occurring during the wet season (June to
140 September) due to monsoons and typhoons, and low inflows in the dry season (December to February) caused by
141 lower temperatures and reduced precipitation. Reservoir releases follow a weekly schedule which is revised by
142 K-water every month based on projections of future storages for the upcoming 3 to 6 months derived using a low
143 inflow scenario (specifically, the 20-year return period drought further described in Section 3.1.1).

144 Figure 2 illustrates the daily reservoir operation records (storages and releases) during the historical drought events
145 analysed in this study. The very severe 2014-16 drought event is included in the analysis for both reservoir systems.
146 This event caused severe damages such as regional water restrictions and nationwide crop failures (K-water, 2018).
147 During this period, the aggregated storage volume for Soyanggang-Chungju reached its lowest in record (1373
148 million m^3, corresponding to 24.3% of storage capacity) and its third lowest record for Andong-Imha (434 million
149 m^3, 23.5%). Additionally, we analysed the drought event from 2001 to 2002 for Soyanggang-Chungju and the
150 event from 2008 to 2009 for Andong-Imha. The drought events show distinct characteristics: both droughts in
151 Soyanggang-Chungju (Figure 2(a, b)) conclude with a large inflow (and outflow) event in the subsequent wet
152 season, while in Andong-Imha relatively low flow conditions persist into the wet season (Figure 2(c, d)).

2.2 Observational data and seasonal weather forecasts

154 This study utilises quality-controlled daily precipitation data from 49 in-situ stations distributed within the
155 catchments, as provided by K-water, along with daily temperature data from 37 in-situ stations managed by the
156 Korean Meteorological Administration (KMA). Unlike precipitation and temperature, potential
157 evapotranspiration (PET) data were computed based on the standardised Penman-Monteith method suggested by
158 the United Nations Food and Agriculture Organisation (Allen et al., 1998). We used the Thiessen polygon method
159 to calculate the mean areal data for each reservoir. For reservoir operation modelling, we used quality-controlled
160 daily reservoir operation records provided by K-water, including the storage volume, inflow and water supply. In
161 generating theses records, K-water utilises a water balance equation, considering reservoir volume changes from
162 water level fluctuations and supplies. These reservoir inflow data are not corrected for removing the effect of
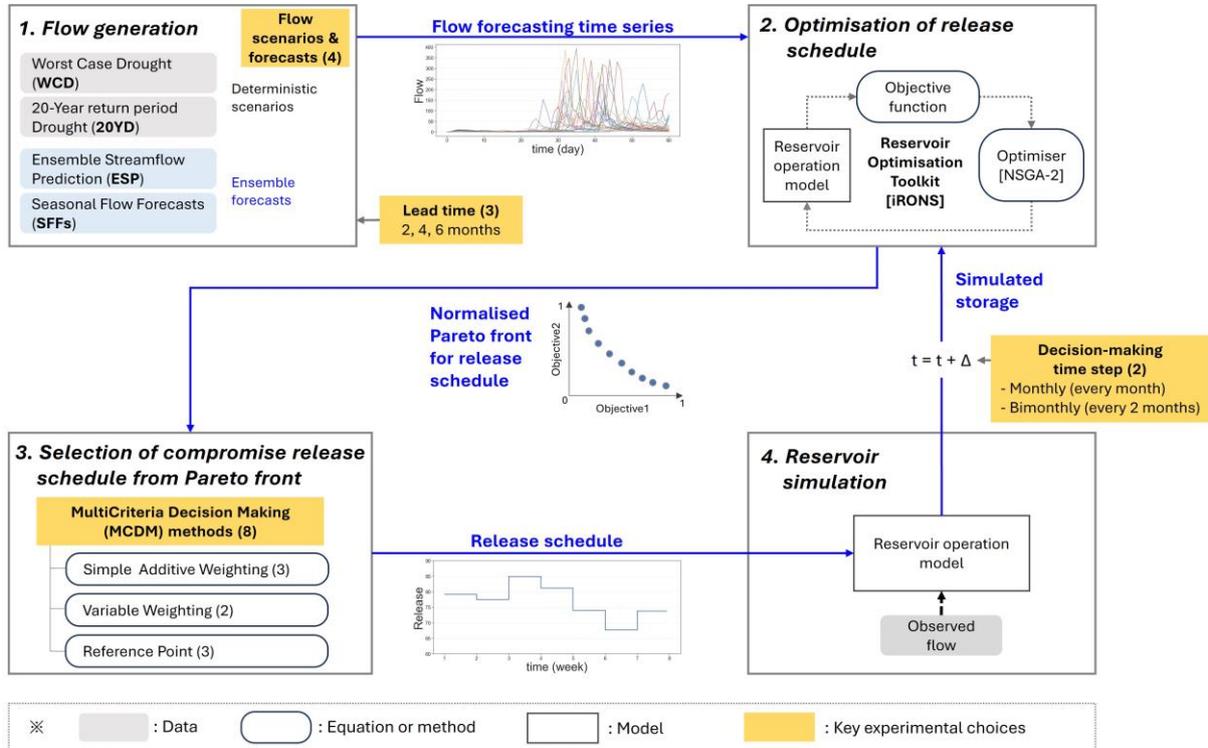163 evaporation losses from the reservoirs.

164 For generating SFFs, we employed the seasonal weather forecasts provided by ECMWF (system 5). This choice
165 was based on our prior research, which demonstrated that generally ECMWF offers the most accurate precipitation
166 forecasts across South Korea (Lee et al., 2023). ECMWF provides 25 ensemble forecasts from 1993 to 2016 and
167 51 ensembles since 2017 on a monthly basis, with a lead time extending up to 7 months ahead. To ensure
168 consistency with our previous works, we obtained ECMWF's seasonal weather forecasts datasets for precipitation,
169 temperature and PET with a spatial resolution of 1×1°. We downloaded forecast data from the Copernicus Climate
170 Data Store every month in the period of the three drought events of Figure 2. Additionally, we downloaded
171 forecast data for the available period from 1993 to 2010 to compute bias correction factors (as detailed in Section
172 3.1.1).

3. Methodology

3.1 Simulating forecasts-informed operations during a past drought event

175 Figure 3 schematically outlines our methodology for simulating reservoir operations during a past drought event,
176 and each compartment of the figure corresponds to a sub-section from 3.1.1 to 3.1.4. We begin by compiling four
177 distinct flow scenarios/forecasts (Worst-Case Drought (WCD), 20-Year return period Drought (20YD) / ESP,
178 SFFs) using historical observational data and seasonal weather forecasts from ECMWF system5 (Section 3.1.1).
179 For each of this flow scenario/forecast, we generate a set of Pareto optimal weekly release schedules, taking into
180 account two conflicting objectives: securing storage volume and minimising supply deficit (Section 3.1.2). A
181 single compromise release schedule within this set is then selected using a Multi-Criteria Decision-Making
182 (MCDM) methods (Section 3.1.3). We then simulate the evolution of the reservoir systems until the next decision-
183 making time step by feeding the chosen release schedule into a reservoir simulation model forced by observed
184 inflow data (Section 3.1.4). The aforementioned process is iteratively repeated until the end of the simulation

185 period (i.e. the end of the historical drought event as shown in Figure 2). As highlighted in Figure 3, the key
186 choices in setting-up each simulation experiment are: the type of flow scenario/forecast and relevant lead time;
187 the MCDM method used to select a compromise schedule; and the decision-making time step, i.e. the frequency
188 with which release schedules are re-optimised. Note that the forecast lead time can be chosen to be longer than
189 the decision-making time step, in which case only the first part of the optimised release schedule is applied before
190 being re-optimised.

191

192 **Figure 3: Schematic diagram illustrating the reservoir simulation methodology employed in this study to simulate**
193 **forecast-informed reservoir operations during a drought event.**

3.1.1   Generation of deterministic flow scenarios and ensemble flow forecasts

195 In this study, we considered two deterministic scenarios, WCD and 20YD, alongside two ensemble forecasts
196 products, ESP and SFFs. All scenarios and forecasts are generated at daily resolution (i.e. the resolution of the
197 hydrometeorological data and weather forecasts) and then aggregated to weekly resolution (the resolution for
198 reservoir simulation). In our simulation experiments, we tested scenarios/forecasts with lead times of 2, 4, or 6
199 months. An example of reservoir simulation process, with various experimental choices such as flow
200 scenario/forecast, lead time and decision-making time step, is illustrated in Figure S2.
201 The WCD scenario was generated by analysing historical flow records and identifying the lowest observed inflows
202 for each reservoir. The 20YD scenario, which is the scenario currently employed in practical reservoir operations
203 in South Korea, was obtained from K-water. To derive this scenario, K-water conducts a low-flow frequency
204 analysis of historical inflow records spanning over 30 years (Ryoo et al., 2009; Jung et al., 2012).
205 We built an ensemble for each weather variable (precipitation, temperature and PET) based on historical
206 observations from 1966 to 2010 and fed it into the Tank hydrological model (Sugawara et al., 1986, 1995) to
207 generate ESP flow ensemble with 45 members (Lee et al., 2024). The Tank model is a lumped conceptual rainfall-
208 runoff model, widely used in South Korea and many other countries (Goodarzi et al., 2020; Ou et al. 2017). We
209 calibrated and validated the model using observations for the period 2001-2010 and 2011-2020, respectively.
210 Lastly, we generated an ensemble of SFFs using ECMWF's seasonal weather forecasts (system 5) as input for the
211 same Tank model (25 ensemble forecasts until 2016 and 51 since 2017). Given the coarse spatial resolution (1°×1°)
212 of the seasonal weather forecast data compared to the reservoir's catchment areas, we applied the linear scaling
213 method to correct biases. Bias correction factors were derived by comparing weather forecasts with observations
214 over the period 1993-2010. We made this choice to maximise the chances of getting robust estimates for the bias
215 correction factors (Maraun et al., 2010; Johnson and Sharma, 2012), although this may lead to an overestimation
216 of the SFFs performance during the 2001-02 and the 2008-09 drought events, given that observations for those
217 events contributed to the bias correction process. Further details regarding the structure, parameter calibration,

5

218 validation and performance of the Tank model, as well as the linear scaling method used for bias correction, are
219 comprehensively documented in our previous paper (Lee et al., 2024).

220 To assess the accuracy of flow forecasts, we employed the Mean Error (ME) of monthly flow averaged across the
221 entire simulation period. In calculating ME for ensemble forecasts, we considered their ensemble median. It is
222 calculated as:

223 $$\text{Mean Error} = \frac{1}{N}\sum_{i=1}^{N}(Q_i^{Forecast} - Q_i^{Observation})$$ (1)

224 where, N represents the total number of timesteps (months) in the simulation periods. $Q_i^{Forecast}$ and $Q_i^{Observation}$
225 are forecasted and observed monthly flow at time i (month), respectively. When the ME is negative (positive),
226 the forecast tends to underestimate (overestimate) the flow.

227 While ME is a simple measure of forecast accuracy, it does not account for the contributions of each member
228 within the ensemble. Therefore, we also computed the forecast skill using the Continuous Ranked Probability
229 Score (CRPS) and the Continuous Ranked Probability Skill Score (CRPSS), developed by Matheson and Winkler
230 (1976). While CRPS measures the absolute performance (score), CRPSS represents the relative performance (skill)
231 with respect to a benchmark, in our case the ESP. These metrics are computed as follows:

232 $$\text{CRPS} = \int [F(x) - H(x \geq y)]^2 \, dx$$ (2)

233 $$\text{CRPSS} = 1 - \frac{\text{CRPS}^{SFFs}}{\text{CRPS}^{ESP}}$$ (3)

234 where $F(x)$ represents the cumulative distribution of the SFFs ensemble, x and y are the forecasted and observed
235 flow. H is called the 'Heaviside (or Indicator) function' and is equal to 1 when $x \geq y$ and 0 when $x < y$. CRPS
236 values range from 0 to infinity and the lower CRPS the higher forecasting performance. CRPS^{SFFs} and CRPS^{ESP}
237 are the CRPS of SFFs and ESP, respectively. When the CRPSS is positive ($0 < \text{CRPSS} \leq 1$), SFFs have skill with
238 respect to ESP, when it is negative, ESP outperforms SFFs. If the CPRSS equals zero, the performance of SFFs
239 is equivalent to that of ESP.

240 To exhibit the skill more intuitively, we employed the concept of 'overall skill', as introduced in our previous
241 research (Lee et al., 2023; 2024). It represents the frequency with which SFFs outperform the benchmark (ESP)
242 over a specific period and can be expressed as:

243 $$\text{Overall skill (\%)} = \frac{\sum_{i=1}^{N} [H(\text{CRPSS})(i)]}{N} \times 100 \, (\%)$$ (4)

244 where N is the total number of months in the periods, i.e. the analysed drought event in our case. Again, the
245 Heaviside function (H) is equal to 1 when CRPSS $(i) > 0$ and 0 when CRPSS $(i) \leq 0$. If the overall skill is greater
246 than 55%, SFFs generally have skill with respect to ESP across the period. However, if it is less than 45%, ESP
247 outperforms SFFs. When the overall skill is between 45% to 55%, we consider them to have an equivalent level
248 of performance (Lee et al., 2024).

3.1.2 Multi-objective optimisation of release schedule

250 Reservoir operations inherently involve managing multiple objectives often in conflict with each other (Zhou et
251 al., 2011; Vassoney et al., 2021). In terms of drought management, the amount of supply deficit shows an inverse
252 correlation with both the secured reservoir storage at the initial stage of the hydrological year (October 1st) and
253 the total inflow into the reservoir across the hydrological year (from October 1st to the subsequent September 30th).
254 In other words, inadequate storage at the outset of the hydrological year leads to substantial disruptions in water
255 supply and the severity of these shortages further increases when the inflow is insufficient. These relationships
256 are evident in historical records, as illustrated in Figure S3.
257 To account for both the need of ensuring supply and of securing storage for the next hydrological year, we
258 established two operational objectives in our optimisation: to minimise the mean Squared Supply Deficit (SSD,
259 [million m^3]^2) over the optimisation period and to minimise the Storage Volume Difference (SVD, million m^3)
260 relative to the reservoir's capacity at the end of the hydrological year. The rationale for squaring the supply deficit
261 is to incorporate risk hedging principles, aimed at strategically allocating water resources over time (You, 2013;
262 Shiau, 2022). These two objectives are formulated as:

263 $$SSD = \frac{1}{T} \sum_{t=0}^{T} [\text{Max}(0, d(t) - Q(t))]^2 \tag{5}$$

264 $$SVD = \text{Max}(0, S_{max} - S) \tag{6}$$

265 where T is the total number of weeks for which the flow forecast is available (i.e. T equals the lead time in months
266 × 4), $d(t)$ and $Q(t)$ represent water demand and supply at week t, respectively. S_{max} is the storage capacity of
267 the reservoir (million m³) and S is the storage volume (million m³) at the end of the hydrological year. When the
268 end of hydrological year is not included in the optimisation period, S is set to the storage at the end of optimisation
269 period. By definition, superior performance is associated with smaller objectives (SSD and SSD).

270 For the reservoir operation modelling and the optimisation of release schedules, this study utilises the 'interactive
271 Reservoir Operations Notebooks and Software' (iRONS) toolbox developed by Peñuela et al. (2021). This toolbox
272 offers a set of Python functions, and a Jupyter Notebook based environment to simulate and optimise reservoir
273 operations. In iRONS, the reservoir model, based on a mass balance equation, is linked to an optimiser that utilises
274 the Non-dominated Sorting Genetic Algorithm (NSGA-II) for the multi-objective optimisation. Given that in
275 multi-objective optimisation problems, a single optimal solution that satisfies all objectives simultaneously is
276 unattainable (Lu et al., 2011; Malekmohammadi et al., 2011), NSGA-II identifies a set of non-dominated solutions
277 whose performance realise different Pareto optimal trade-offs between the two objectives. The performances
278 associated with these solutions visualised in the objective space constitute the so called "Pareto front" (Giagkiozis
279 and Fleming, 2014; Ni et al., 2022). The SSD and SVD are used as objective functions to generate a Pareto front.
280 We set the number of solutions to be evolved by the NGSA-II algorithm (so called "population" size) to 100, and
281 the number of iterations to 100000, leading to a total of ten million model evaluations for each optimisation run.
282 When optimising against ensemble forecasts, the two objective functions (Eqs. 5 and 6) are evaluated against each
283 ensemble member, and the average is taken as the final objective value and passed on to the NSGA-II optimiser.

284 ### 3.1.3   Selection of a compromise solution from the Pareto front

285 Since the Pareto front delivered by the multi-objective optimisation (Section 3.1.2), comprises multiple release
286 schedules, a critical decision must be made to select one compromise release schedule from that Pareto front. The
287 methodology for this selection will be referred to as Multi-Criteria Decision-Making (MCDM) from now on, as
288 described in some of the previous literature (e.g. Wang and Rangaiah, 2017; Ni et al., 2022). In other context,
289 MCDM methods are presented to support decision makers in selecting compromise alternatives for complex water
290 management issues (Afshar et al., 2011; Malekmohammadi et al., 2011; Zhu et al., 2017; Vassoney et al., 2021).
291 In this study, the MCDM method is employed as a way to mimic the selection that, in real world, would be made
292 by the reservoir operator when running forecasts through a reservoir operation optimisation model and being
293 returned a Pareto front. Given the significant uncertainty regarding how the operator would make this selection,
294 considering multiple MCDM methods provides a means to address this uncertainty in our assessment of forecast
295 value.

296 Various MCDM methods have been developed and utilised over the last several decades (Velasquez and Hester,
297 2013). Among them, this study employed eight distinct methods, which can be systematically categorised into
298 three groups: Simple Additive Weighting (SAW), variable weighting and reference point methods. Firstly, the
299 SAW method, which is frequently employed in decision-making (Arsyah et al., 2021), ranks the alternatives based
300 on their weighted sum performance (Fishburn, 1967). In this study, we consider the 'balanced' method where
301 equal weights are assigned to each objective, as well as the 'storage-prioritised' and 'supply-prioritised' methods,
302 which prioritise storage and supply, respectively.
303 Secondly, we propose the 'variable weighting' method, which reproduces more closely the thought process of
304 reservoir operators, who weight supply more when the storage is abundant and less when storage is scarce. We
305 applied this method in two ways: the 'simple selective' method, which adopts the same weights as in the SAW
306 methods but varying them depending on storage status and the 'multi-weight' method, which applies more detailed
307 procedure to allocate weights based on storage status.
308 Lastly, the reference point method identifies the compromise solution on a Pareto front by measuring the distance
309 from a reference point. In this study, we applied three approaches: the 'utopian point', 'knee point' and 'TOPSIS'
310 methods. The utopian point method selects the solution on the Pareto front that minimises the Euclidean distance
311 from the utopian (or ideal) point, which represents the theoretical perfect solution (Lu et al., 2011). The knee point
312 method selects the knee point, which is a point where the curvature of the Pareto front is maximum (Das, 1999).
313 Among various methods for detecting the knee point, we employed the Minimum Manhattan Distance method
314 which is known for its simplicity and robustness (Chiu et al., 2016; Li et al., 2020). The TOPSIS method selects
315 a point with the shortest Euclidean distance from the ideal point and the longest distance from the anti-ideal point
316 as the compromise solution (Hwang and Yoon, 1981; Liu, 2009). This is a widely chosen method (Tzeng and

317 Huang, 2011; Wang and Rangaiah, 2017) including the United Nations Environmental Program (Chen, 2000; Zhu
318 et al., 2015).
319 Detailed information on the MCDM method and normalisation of a Pareto front, including equations, merits and
320 demerits, is provided in the supplementary material (Section S1 and S2).

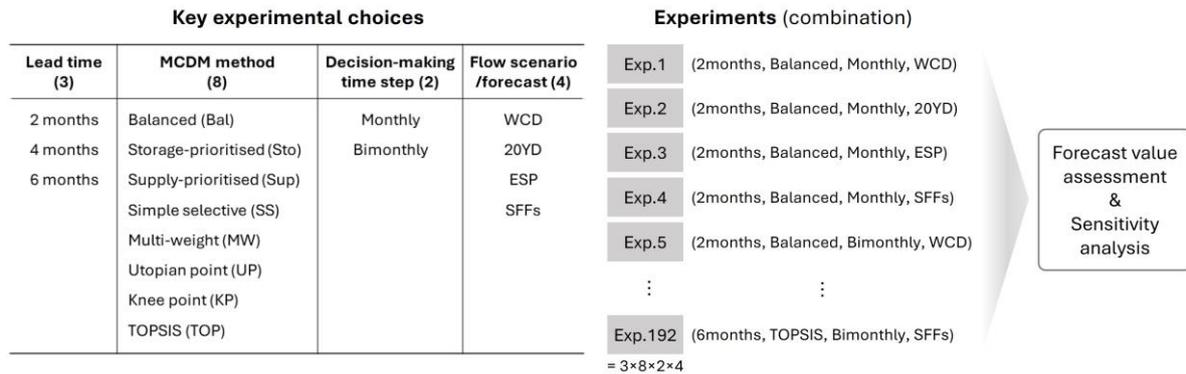### 3.1.4 Reservoir simulation against observed inflows

322 Once a Pareto-optimal release schedule is selected, the final step of our simulation methodology is to evaluate
323 what would have been achieved if that schedule had been implemented. To this end, we simulate how the reservoir
324 storage would have evolved under the selected release schedule and against observed inflows. The reservoir
325 simulation is based on the repeated application of the following water balance equation:

326
$$S_{t+1} = S_t + I_t - R_t \qquad\qquad t = 0, \ldots, \text{T-1} \tag{7}$$

327 where S_t is the simulated storage, I_t the observed inflow and R_t the optimised release for week t. Eq. 7 is repeated
328 from time t=0, corresponding to the week when the scenarios/ forecasts are generated and the optimisation is run,
329 until the time when the generation/optimisation process is run again, i.e. time T=4 (weeks) in the case of monthly
330 decision-making time-step or T=8 (weeks) in case of bimonthly. The final simulated storage S_T is then used as
331 the initial storage for the next multi-objective optimisation run and subsequent simulation (see Figure 3).

3.2 Measuring the forecast value and its sensitivity to experimental choices

333 For each drought event, the reservoir simulation of the forecast-informed operations described in Section 3.1, is
334 repeated using different scenario/forecast products and with various combinations of key experimental choices.
335 These choices include the forecast lead time, the MCDM method and the decision-making time step, as
336 summarised in Figure 4. Therefore, the total number of experiments for each drought event amounts to 192 (3
337 lead times × 8 MCDM methods × 2 decision-making time steps × 4 flow scenarios/forecasts).

※ **MCDM**: MultiCriteria Decision-Making
WCD: Worst Case Drought, **20YD**: 20-Year return period Drought, **ESP**: Ensemble Streamflow Prediction, **SFFs**: Seasonal Flow Forecasts

338

339 **Figure 4: Key experimental choices for simulating forecast-informed reservoir operations. Each of the 192 simulation**
340 **experiments is conducted according to Figure 3.**

341 For each experiment, we computed two performance indicators, representing supply deficit (SSD) and storage
342 volume (SVD) as in Eqs. 5 and 6 but using the simulated storage and release time series from the simulation
343 against observed inflows (step 4 in Figure 3). We then calculated the same indicators using the observed storage
344 and release, to quantify the performance of the historical operations, which we use as a benchmark. Unlike
345 previous studies (e.g. Turner et al., 2017; Peñuela et al., 2020a; Crippa et al., 2023) that analysed improvements
346 in performance indicators separately, here we propose a new and simple way to take into account the improvement
347 in both indicators simultaneously. In fact, performance indicators generally exhibit a trade-off relationship with
348 each other, so that an improvement with respect to the benchmark for one indicator may come at the price of a
349 loss in the other. Analysing them independently from one another obfuscates these trade-offs.
350 To overcome this issue, here we calculated the difference in each indicator (simulated - historical) in each
351 experiment and defined the forecast value as the number of experiments where this difference is negative for both
352 indicators. In fact, since we aim to minimise both indicators, negative differences in both indicate that the
353 simulated operations outperform the historical operation. This method provides an intuitive and practical
354 understanding of forecast value, as it directly relates to the chances of achieving better operational outcomes
355 compared to historical operation, taking into account operational trade-offs and factoring in the uncertainty in the
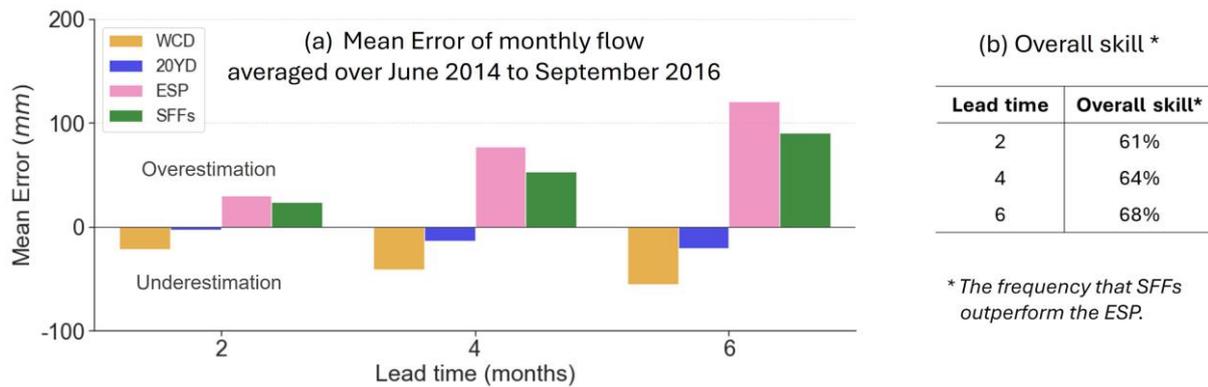
8

356 key experimental set-up choices. Last, we analysed the sensitivity of forecast value to those key experimental
357 choices. This analysis serves as a useful tool in pinpointing the primary determinant of forecast value and offering
358 insights for optimising setup choices to maximise the value for drought management.

4. Results

360 For clarity of illustration, in Section 4.1, we first present the results for one event and reservoir system: the drought
361 that occurred in Soyanggang-Chungju from 2014 to 2016 (Figure 2). In Section 4.2, we expand our results to
362 include other reservoir systems and events, aiming to explore to what extent our conclusions on the value of SFFs
363 and its key controls can be generalised.

4.1 Simulation results for the 2014-2016 drought in Soyanggang-Chungju reservoirs

4.1.1   Accuracy and skill of seasonal flow forecasts

366
367 **Figure 5: (a) Mean Error of monthly flow (simulated – observed) for Soyanggang-Chungju reservoir system averaged**
368 **from June 2014 to September 2016 for different scenarios/forecasts (2-month lead time). (b) The overall skill of SFFs,**
369 **which represents the frequency of SFFs outperforming ESP across the simulation period.**

370 Figure 5(a) illustrates the Mean Error of monthly flow (see Eq. 1) for lead times of 2, 4, 6 months and different
371 type of flow scenario/forecast (WCD: yellow, 20YD: blue, ESP: pink, SFFs: green). As shown in the figure,
372 deterministic scenarios (WCD and 20YD) exhibit smaller errors compared to the ensemble forecasts. This is not
373 surprising, as the WCD and 20YD scenarios are designed to mimic dry conditions and we are now evaluating
374 accuracy on a severe drought event. Ensemble forecasts, particularly, show a systematic bias towards
375 overestimating flows, with this tendency being more pronounced in ESP compared to SFFs. This pattern is
376 consistently observed across different reservoir systems and events, as further illustrated in Figure S4.
377 Figure 5(b) shows the overall skill (see Eq. 4), indicating the frequency with which SFFs outperform ESP across
378 the simulation period. In this specific event, the overall skill exceeds 60% at all lead times, indicating that SFFs
379 generally perform better compared to ESP. However, results from different reservoir systems and events (reported
380 in Figure S4), show lower overall skill, and decreasing with lead time. Our additional analysis of cumulative flow
381 observations and forecasts for the period 2014-16, presented in Figure S5, indicates that this drought event was
382 more severe than the 20-year return period drought for nearly two years – until the high inflows of July 2016.
383 Given that reservoirs in South Korea are designed to supply water for a year under a drought with a 20-year return
384 period, this event posed significant challenges to reservoir operators.

4.1.2   Reservoir simulations and their performances

386 The simulated reservoir operation results are illustrated in Figure 6, showing the storage volume (a) and
387 cumulative squared supply deficit (b) generated using WCD (yellow), 20YD (blue), ESP (pink) and SFFs (green).
388 For each flow scenario/forecast, there are 48 simulation outcomes resulting from different combinations of the
389 experimental choices (3 lead times × 8 MCDM methods × 2 decision-making time steps). Higher storage volume
390 compared to historical operation (black line) is preferable and vice versa for cumulative squared supply deficit.

391

Figure 6: Simulated reservoir operation results for Soyanggang-Chungju from June 2014 to September 2016 in terms of (a) storage volume and (b) cumulative squared supply deficit. From top to bottom, the rows represent simulation of the forecast-informed operations using WCD (orange), 20YD (blue), ESP (pink) and SFFs (green), respectively. Each sub-figure has 48 simulated results (coloured lines, 3 lead times × 8 MCDM methods × 2 decision-making time steps) and a single historical operation (black line). The numbers on the right end of Figure 6(a) represent the mean storage volume (million m³) across all 48 simulations at the end of the simulation period (September 30th, 2016).

As shown by the black lines in Figure 6, the reservoir operators seemed to be unaware of the impending drought event until June 2015, as up to that point they continued to supply the demand (causing no deficit) while storage level declined. Subsequently, their operational focus shifted to managing storage availability, which led to significant supply deficits beginning in July 2015. Compared to this historical operations, most of the forecast-informed reservoir operations achieve higher storage volumes throughout the event (Figure 6(a)). By the end of the simulation period (September 30th, 2016), all forecast-informed operations replenish the reservoir system more than the historical operations did. Operations informed by deterministic scenarios (upper two rows) offer slightly superior results for securing storage volume compared to ensemble forecasts (lower two rows), as shown by average final storage values reported at the right end of Figure 6(a), but they produce larger supply deficits than ensemble forecasts (Figure 6(b)). This trend arises from the underestimation of flows by deterministic scenarios (see Figure 5(a)), which results in reduced releases and increased supply deficits. (Further results for other reservoir systems and drought events also depicted in Figure S6.)

In particular, many of the ensemble members of SFFs produced in June 2016 anticipated the high flow event that occurred in July 2016 (see Figure S7). This led the multi-objective optimisation informed by SFFs to suggest higher releases, in contrast to the (unnecessarily) low releases designed when using the worst case or 20-year return period drought scenario.

4.1.3 Value of seasonal flow forecasts

Figure 7 depicts the differences in achieved performance indicators (SSD and SVD) between simulated operations and historical operation at distinct points in time (X, Y and Z in Figure 6), corresponding to the end of hydrological years (September 30th). Coloured circles in the figure denote the type of flow scenario/forecast used in simulations (following the same colour coding as in Figure 6) and there are 48 circles (3×8×2) in each colour, corresponding to the combinations of 3 lead times, 8 MCDM methods and 2 decision-making time steps. Circles positioned below (above) zero for both the x and y axes, i.e. within the green (red) shaded area, indicate experiments where reservoir simulations achieve better (worse) performance compared to historical operation. The count of circles within the green shaded area (bottom-left quadrant) represents the forecast value, indicating the chances of simulated reservoir operations outperforming historical operation, as detailed in Section 3.2.
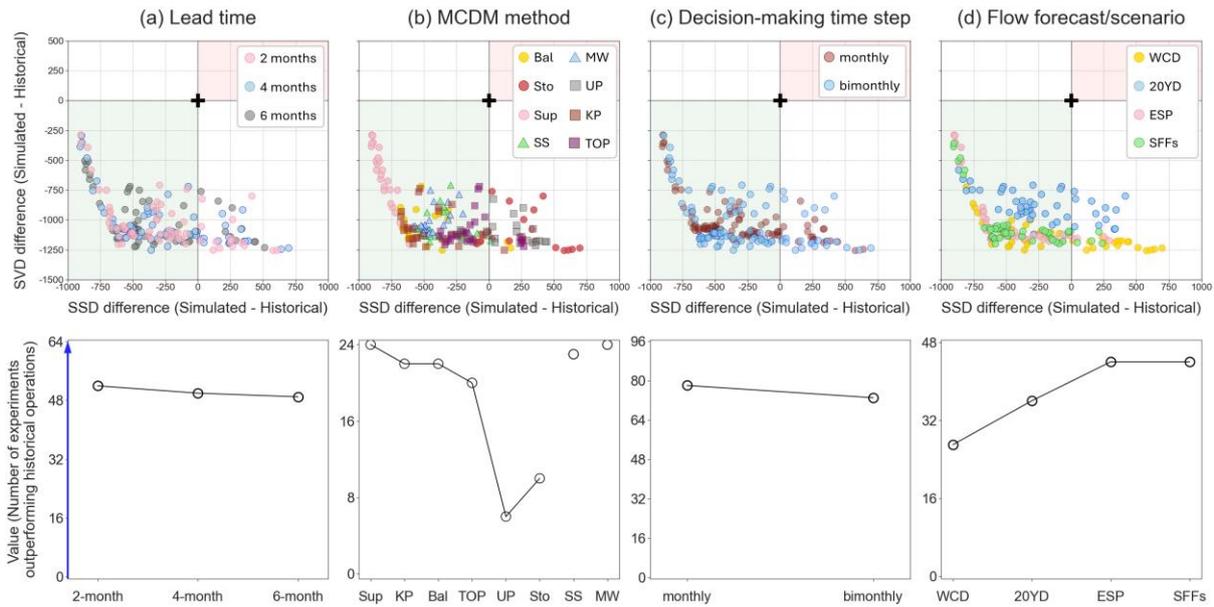
10

Figure 7: Difference in SSD (x-axis) and SVD (y-axis) between historical operation (black cross) and simulated operations using different flow scenarios/forecasts (coloured circles) in Soyanggang-Chungju during the 2014-2016 drought. Performances are calculated on September 30th in (a) 2014, (b) 2015 and (c) 2016. Each sub-figure shows 48 points for each flow scenario/forecast (WCD, 20YD, ESP, SFFs), resulting from different combinations of key experimental choices (3 lead times × 8 MCDM methods × 2 decision-making time steps).

At the initial stage of simulation, as shown in Figure 7(a), simulated forecast-informed operations only exhibit a trade-off relationship with historical operation. All circles are distributed in the bottom-right quadrant, indicating that the historical operation prioritised water supply over storage volume until the end of September 2014 (X). However, as the impact of forecast-informed operations accumulates (i.e. the period of simulation moves from X to Z), more circles tend to fall in the green shaded area where simulated operations outperform historical operation. This result suggests that the model-based reservoir operation optimisation has the potential to improve the management of prolonged drought events. Specifically, as shown in Figure 7(c), the majority of simulations not falling within the green shaded area by the end of the simulation (September 30th, 2016 (Z)), are associated with deterministic scenarios (yellow and blue circles). These findings are consistently demonstrated with our experiments applied to other drought events and reservoir systems, as presented in Figure S8.

4.1.4 Sensitivity of forecast value to key experimental choices

The top row of Figure 8 presents a figure similar to Figure 7(c), but with distinct colour codes assigned to different experimental choices for each category. The bottom row of Figure 8 illustrates the sensitivity of forecast value (y-axis) to the choice of forecast lead time (a), MCDM method (b), decision-making time step (c) and type of flow scenario/forecast (d) for September 30th, 2016 (corresponding to Figure 7(c)). The maximum number on the y-axis in each sub-figure represents the total number of simulation experiments conducted for a particular experimental set-up choice. The forecast value (hollow circle) represents the number of experiments that the reservoir simulation outperforms historical operation for both objectives (SSD and SVD). For example, in the bottom row of Figure 8(a), the lead time is fixed at 2, 4 or 6 months (horizontal axis) and for each of these choices there are 64 experiments (see range of vertical axis), resulting from the combination of 8 MCDM methods, 2 decision-making time steps and 4 flow scenarios/forecasts. When an experimental choice (x-axis) correlates with a higher forecast value (y-axis), it indicates that using that specific experimental choice can lead to greater operational benefits for managing droughts.

453

Figure 8: Top row: Difference in SSD (x-axis) and SVD (y-axis) between historical operation (black cross) and simulated operations using different experimental choices: (a) forecast lead time, (b) MCDM method, (c) decision-making time step and (d) type of flow scenario/forecast. Bottom row: Forecast value (y-axis) plotted against the same experimental choices for the same reservoir system and date. The MCDM methods are ordered from left to right with increasing importance to storage availability, along with two variable weighting methods (SS and MW). All results refer to Soyanggang-Chungju reservoir system on September 30th, 2016.

As shown in Figure 8(b), the number of experiments outperforming historical operations varies significantly depending on the MCDM method used for selecting a compromise solution from the Pareto front, suggesting that this choice is a key control of forecast value. In this specific drought event, using the storage-prioritised (Sto) and utopian point (UP) methods leads to a much lower forecast value compared to using the other methods. Note that, with our problem formulation, the storage-prioritised and utopian point methods are the ones that give more weight to conservation of storage volumes, at the expenses of supply deficits (see Figure S9 for further details). Importantly, Figure 8(d) demonstrates that the value is also be influenced by the type of flow scenario/forecast used to inform the reservoir operations optimisation. In this case, a higher value is attained using ensemble forecasts (ESP, SFFs) than deterministic scenarios (WCD, 20YD), but there is no difference in forecast value between ESP and SFFs.

Additionally, we applied a bootstrapping technique to test the impact of using different sample sizes across the plots in Figure 8 (bottom row) and found that the impact of sample sizes on sensitivity result is negligible (see Figure S10).

4.2 Simulation results for other reservoir systems and drought events

4.2.1 Sensitivity of forecast value to key experimental choices

Having analysed the forecast value and its key controls for one drought event in one reservoir system, Figure 9 illustrates whether similar or contrasting results are found in the other three events and reservoir systems considered in this study (see Figure 2 for a description of these events; intermediate results, i.e. simulated storages and releases, and difference in both objectives for these reservoir systems and events are reported in Figure S6 and Figure S8). Note that Figure 9 incorporates the result from Soyanggang-Chungju for the 2014-2016 drought event already shown in Figure 8 (white circles connected by solid line).

12

481

Figure 9: Forecast value (y-axis) against key experimental choices including (a) lead time, (b) MCDM method, (c) decision-making time step and (d) type of flow scenario/forecast for Soyanggang-Chungju (○, □) and Andong-Imha (▲,▼) at the end of different drought events (points Z, K, M in Figure 2). The MCDM methods are ordered the same as in Figure 8.
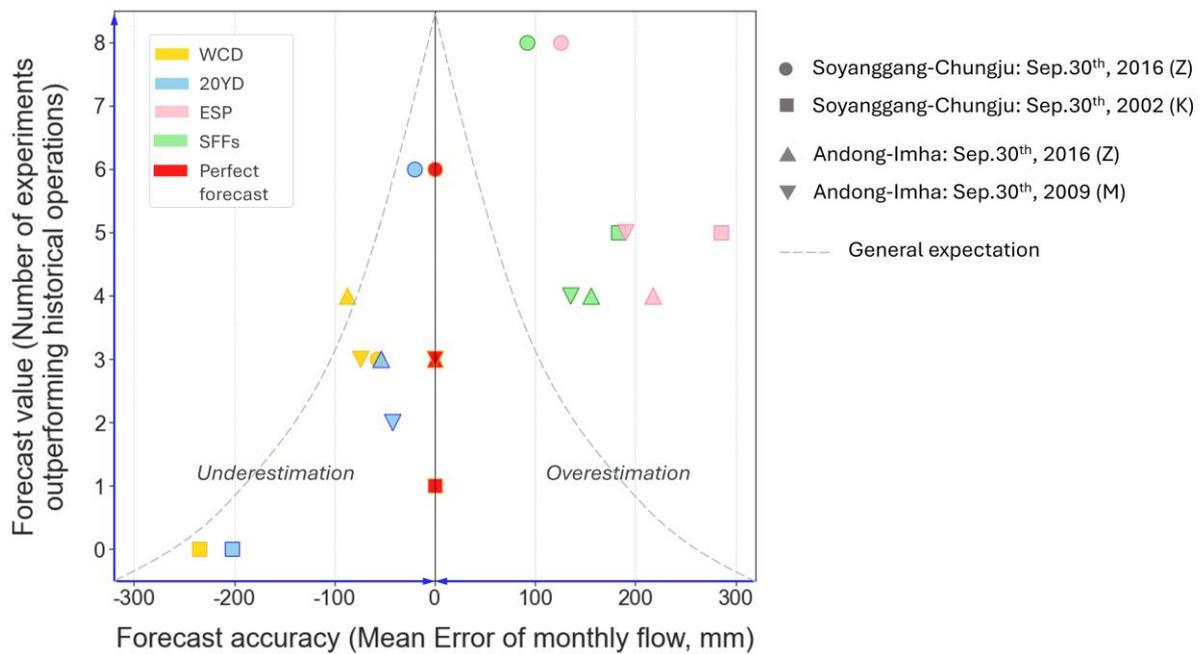
Figure 9(b) confirms the substantial influence of the choice of MCDM method on forecast value. However, it also highlights that which method deliver more values varies with the event and reservoir system. As already noted, in the Soyanggang-Chungju reservoir system, the forecast value increases with MCDM methods that prioritise avoiding supply deficits (i.e. Sup, KP and Bal). This is likely due to the fact that the two drought events analysed for this system end with a large inflow event (see Figure 2), therefore the emphasis on minimising supply deficits, combined with the forecast ability to anticipate the upcoming inflows (as discussed in Sec. 4.1.2), enable to fully exploit the natural replenishment of storage from the wet event that occurs at the end of the simulation period. The opposite is observed in Andong-Imha system, where the analysed drought events persist into the upcoming wet season, and therefore MCDM prioritising storage conservation (UP, Sto) tend to deliver more value. The higher value of ensemble forecasts (ESP and SFFs) is also confirmed in Figure 9(d), particularly in Soyanggang-Chungju reservoir system, whereas their advantage over deterministic scenarios (WCD and 20YD) is less pronounced in Andong-Imha. Lastly, Figure 9(a,c) indicate that increasing the forecast lead time or decreasing the decision-making time step slightly improves forecast value. Yet again, this improvement appears relatively marginal when compared to the impact of the chosen MCDM method or flow scenario/forecast.

4.2.2   Relationship between forecast accuracy and value

Figure 10 illustrates the overall relationship between the accuracy of each flow scenario/forecast (x-axis) and its value computed over the 8 MCDM methods (y-axis) in informing decision-making for enhanced drought management. For this figure, we only used experiments with a 6-month lead time and monthly decision-making, to closely mimic current reservoir operations practices in South Korea. Note that excluding other options for these two experimental choices should not undermine the robustness of our conclusions as the sensitivity analyses in previous sections have shown that these choices have low impact on forecast value. Red symbols represent the simulation results when using observations of future flows as if they were 'perfect' forecast (note that, by construction, this scenario is associated with zero error on the horizontal axis of Figure 10).

13

509

Figure 10: Relationship between forecast accuracy (Mean Error of monthly flow, x-axis) and value (calculated over the 8 MCDM methods, y-axis) at the end of the simulation period for different drought events and reservoir systems. For each event and system, the figure shows five points corresponding to the simulated forecast-informed operations using different scenarios/forecasts (orange: WCD, blue: 20YD, pink: ESP, green: SFFs, red: perfect forecast). The perfect forecast scenario was generated using actual flow observations as future forecasts. The direction of the blue arrows indicates higher performance (high value, low error). The grey dashed lines conceptually illustrate the general expectation on the relationship between forecast accuracy and value.

Figure 10 demonstrates that despite the higher accuracy of deterministic scenarios, as evidenced by general proximity of the yellow and blue points to zero on the x-axis, ensemble forecasts (pink and green points) result in a higher value. The Pearson's correlation coefficient between the accuracy and value for the datapoints of Figure 10 is approximately -0.2, indicating a very weak relationship. These findings deviate somewhat from the general expectation on the relationship between forecast accuracy and value that higher accuracy would lead to higher value (this assumed relationship is represented in Figure 10 by the dashed grey lines). When comparing the accuracy between ensemble forecasts, SFFs demonstrate a slight advantage over ESP, with a tendency for smaller overestimations. In terms of forecast value, however, there are no significant differences between them, indicating the operational benefits obtained from using SFFs and ESP appear to be comparable.

Figure 10 also includes the value obtained from optimising operations against the perfect forecast scenario, depicted by red symbols. Surprisingly, this figure indicates that the value of perfect forecast in our experiments is lower than that of ESP and SFFs. This counterintuitive result stems from the fact that even with perfect knowledge of flows within the optimisation horizon (i.e. the forecast lead time), perfect forecast does not resolve the uncertainty about future flows beyond that horizon. Therefore, acknowledging uncertainty during the optimisation horizon, as done when using ensemble forecasts, yields more cautious operations that in the long-term prove to be more robust against adverse events not seen during the optimisation.

5. Discussion

5.1 Value of SFFs in informing decision-making for managing droughts

Our findings highlight the higher value of ensemble forecasts (ESP, SFFs) over deterministic scenarios (WCD, 20YD), aligning with several previous studies. For example, Peñuela et al. (2020a) demonstrated that employing ensemble forecasts can yield higher operational benefits compared to using deterministic (worst-case) scenario in a water supply reservoir system in the UK. The higher value of ensemble forecasts for informing flood control decisions was also demonstrated by Fan et al. (2016). They compared the value using the ensemble mean versus using the full SFFs ensemble and found that the latter notably enhanced forecast value. However, our research also revealed that the extent to which ensemble forecasts yield higher value can vary significantly depending on the reservoir systems, as the enhancement of operational benefits was more evident in the Soyanggang-Chungju than in the Andong-Imha reservoir system. It is also notable that even a perfect forecast with zero forecasting

14

544 error did not achieve a higher value compared to the ensemble forecasts (ESP and SFFs). The lower performance
545 of the perfect forecast scenario is counterintuitive but can be attributed to its finite lead time. In other words,
546 accounting for uncertainty within the optimisation horizon, as done by ESP and SFFs, indirectly helps to better
547 handling the uncertainty about inflows beyond that horizon (refer to Figure S11). Few previous studies also
548 reported that forecast-informed operations forced by ensemble forecasts often deliver comparable or higher
549 performance compared to the perfect forecast scenario (Zhao et al., 2011; Fan et al., 2016; Ficchì et al., 2016).
550 While our findings emphasise the importance of considering forecast uncertainty when optimising reservoir
551 operations, no significant difference in value was found between the two ensemble forecasts (ESP and SFFs). This
552 is consistent with the findings of Peñuela et al. (2020a), who similarly observed no notable difference in the value
553 of ESP and SFFs. Given the lower computational cost and higher practical experiences of generating ESP, the
554 latter remains a hard-to-beat reference.

555 For analysing the relationship between forecast performance and value, we only evaluated two attributes of the
556 forecast performance (i.e. accuracy and skill). Our results showed that the relationship between the forecast
557 performance and value is not significant. To further explore this relationship, additional attributes, such as
558 correlation, variance and reliability may also be considered. These attributes might yield different outcomes when
559 comparing forecast products and could provide new insights into the relationship between performance (i.e., the
560 level of agreement between forecasts and observations) and value (i.e., their usefulness in informing decisions).

561 This study includes a sensitivity analysis that examines how forecast lead time, MCDM method, decision-making
562 time step and type of flow scenario/forecast affect value. Although we found some improvements in forecast value
563 with longer lead times, their impact was generally marginal. A prior study by Yang et al. (2021) also evaluated
564 the influence of lead time, ranging from 10 to 30 days, on forecast value for hydropower and water supply. They
565 argued that considering a longer lead time for forecast-informed operations may enhance the value. However, the
566 lead times they examined were considerably shorter than those in our study, which makes direct comparisons with
567 our study challenging. To further validate the relationship between lead time and forecast value, it is essential to
568 conduct additional research involving a broader range of reservoirs and drought events.

569 The highly variable performance of MCDM methods depending on reservoir systems and drought events (see
570 Figure 9(b)) emphasises the significance of using ensemble forecasts in reservoir operations, as they consistently
571 bring operational benefits (see Figure 9(d)). While identifying an optimal MCDM method which could offer the
572 best solution across all drought events was not possible in this study, practical guidelines can be offered for
573 applying each method based on their inherent characteristics. Firstly, the SAW method is straightforward to apply
574 and may be particularly advantageous for reservoirs with obvious operational purposes or characteristics.
575 Specifically, the supply-prioritised method might be well-suited for a reservoir with ample storage capacity but
576 lower demand. On the other hand, the storage-prioritised method would be useful for reservoirs with a high risk
577 of causing significant economic or social damage when facing a substantial supply deficit over short periods. This
578 method helps mitigate the risk of extreme storage shortages, thereby reducing the likelihood of accidental supply
579 failure (e.g. zero supply for certain periods). Secondly, the performance of the variable weighting method can be
580 highly dependent on subjective choices in determining the appropriate weights and storage ranges. Therefore,
581 sufficient operational records are essential for effectively applying this method. Conversely, the reference point
582 method, offering a geometric estimation of the compromise solution, may prove advantageous for reservoirs with
583 limited operational history.

584 Bias correction of seasonal weather forecasts, such as precipitation, is a widely addressed issue concerning the
585 performance of SFFs (Shrestha et al., 2017). In this study, we utilised bias-corrected SFFs, building on our
586 previous findings that demonstrated the effectiveness of bias correction in improving the accuracy of SFFs (Lee
587 at al., 2023b). While the positive impact of bias correction on SFFs is widely documented in the literature (e.g.
588 Lucatero et al., 2018; Tian et al., 2018; Pechlivanidis et al., 2020), a previous study noted that bias correction may
589 potentially reduce performance under extreme conditions (Crochemore et al., 2016). Our supplementary
590 experiment, presented in Figure S12, investigates the influence of bias correction on forecast value. The result
591 indicates that bias-corrected SFFs generally yield higher value compared to SFFs without bias correction.
592 However, to fully validate the impact of bias correction on the value, further research applying our methodology
593 across diverse reservoirs and drought events is necessary.

5.2 Limitations and directions for future research

595 A key limitation of our study is the limited range of drought events analysed. While we assessed forecast value
596 across two reservoir systems and three historical drought events, these samples are still limited to draw general
597 conclusions. This limitation is difficult to overcome given the infrequent occurrence of extreme drought events

and the limited availability of seasonal forecast data, which only became available in 1993. Since our results have shown the dependency of forecast value on reservoir systems and events, more assessments are needed to establish more general patterns in the relationship between accuracy and value, as well as to compare the performance between different forecast products. We hope that the methodology and open-source code developed for this study will enable potential users to replicate our experiments and validate our provisional results across other regions around the world.

Secondly, in assessing forecast value, we use historical operational performance as a benchmark. While it offers more intuitive value comparison, it is important to recognise that historical operations may have been influenced by a range of internal and external circumstances not captured by our model and performance indicators. For example, reservoir release decisions may be adjusted based on additional water supplies from external sources such as neighboring reservoirs or rivers. Additionally, our proposed method only looks at whether the historical performance is Pareto-dominated, but it does not account for the magnitude in differences between historical and simulated performances. Incorporating hypervolume, defined as the space enclosed by a set of points in a multi-dimensional space (While et al., 2006; Sanchez-Gomez et al., 2019), could enhance this method to better quantify the value.

Last, our modelling of the reservoir systems is based on several simplifying assumptions. A key simplification is that evaporation from reservoirs is not considered. In South Korea, direct measurements of resesrvoir evaporation are rarely conducted, which poses challenges to ensuring the reliability of indirect evaporation estimation. Recent research by Park et al. (2024) introduced an empirical formula to estimate reservoir evaporation specifically for Yongdam Reservoir, which is uniquely equipped with direct evaporation measurements. The study highlighted the importance of validating this formula for its applicability to other reservoirs. However, reservoir evaporation tends to intensify during extreme droughts, resulting in increased loss of storage volume (Wurbs and Ayala, 2014; Shah et al., 2024). Thus, further studies incorporating reservoir evaporation based on reliable estimating formulas are necessary. A second simplication is that we only used two operational objectives: securing storage and minimising water deficit, to centre our attention on reservoir operations. However, there are various other objectives worth considering for simulating reservoir operations, such as potential economic damages from droughts or benefits of risk hedging. Although quantifying those objectives is challenging, incorporating them into a multi-objective approach for drought management could significantly assist water managers.

6. Conclusions

This study explores the potential usefulness of SFFs in informing reservoir operations for managing droughts in South Korea. While deterministic scenarios (WCD, 20YD) exhibited higher accuracy, the value achieved from using ensemble forecasts (ESP, SFFs) was higher. This result emphasises the significance of considering flow forecast uncertainty when optimising reservoir operations and demonstrates that higher forecast accuracy does not necessarily translate into higher value. Our study also suggests that forecast-informed operations using ensemble forecasts can reduce supply deficit and increase storage conservation compared to historical operations during past drought events. However, no clear evidence was found supporting that SFFs can lead to greater value over conventional ESP at present. As seasonal weather and flow forecasting technology continuously evolves and improves, this conclusion is provisional, and it will be important to continue to assess the performance of SFFs in enhancing reservoir operations as new forecasting products become available. Our sensitivity analysis also shows that the MCDM method used to select a compromise release schedule from a Pareto front is a key control of forecast value. This suggests that the operator's prioritisation of competing objectives is crucial in determining forecast value.

By analysing multiple reservoirs and drought events within the same region, our study takes an initial step toward systematising the forecasts performance and value assessment. While this effort is still incomplete, it serves as a beginning to move beyond the "single case study" approach that has dominated previous research in this area. We hope that the workflow and open-source code developed in this study will help researchers and water managers in South Korea as well as other countries in conducting further research and expanding the practical application of SFFs to enhance drought management. In particular, we proposed a new simple method to assess the forecast value that simultaneously takes into account the trade-offs between operational objectives and the uncertainty stemming from key set-up choices for the simulation experiments. This is achieved by counting the number of simulation experiments that outperform benchmark operations (the historical operations in our case) for both objectives. This straightforward performance metric may be useful for quantifying forecast value in a practical and intuitive manner across a wide range of water resources management studies, beyond drought management, including hydropower, flood control and other applications.

References

Afshar, A., Mariño, M.A., Saadatpour, M. and Afshar, A.: Fuzzy TOPSIS multi-criteria decision analysis applied to Karun reservoirs system, *Water Resources Management*, 25(2), 545–563, https://doi.org/10.1007/s11269-010-9713-x, 2011.

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M.: *Crop evapotranspiration: Guidelines for computing crop water requirements*, United Nations Food and Agriculture Organization, Irrigation and drainage paper 56, Rome, Italy, 1998.

Alley, R.B., Emanuel, K.A. and Zhang, F.: Advances in weather prediction. *Science*, [online] 363, 342–344, https://doi.org/10.1126/science.aav7274, 2019.

Arnal, L., Cloke, H.L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B. and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, *Hydrology and Earth System Sciences*, 22(4), 2057–2072, https://doi.org/10.5194/hess-22-2057-2018, 2018.

Arsyah, U.I., Jalinus, N., Syahril, Ambiyar, Arsyah, R.H and Pratiwi, M.: Analysis of the Simple Additive Weighting method in educational aid decision making, *Turkish Journal of Computer and Mathematics Education*, 12(14), 2389–2396, https://doi.org/10.17762/turcomat.v12i14.10664, 2021.

Baker, S.A., Rajagopalan, B. and Wood, A.W.: Enhancing ensemble seasonal streamflow forecasts in the upper Colorado river basin using multi-model climate forecasts, *Journal of the American Water Resources Association*, 57, 906–922, https://doi.org/10.1111/1752-1688.12960, 2021.

Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55, https://doi.org/10.1038/nature14956, 2015.

Block, P.: Tailoring seasonal climate forecasts for hydropower operations, *Hydrology and Earth System Sciences*, 15(4), 1355–1368, https://doi.org/10.5194/hess-15-1355-2011, 2011.

Chen, C.-T.: Extensions of the TOPSIS for group decision-making under fuzzy environment, *Fuzzy Sets and Systems*, [online] 114(1), 1–9, https://doi.org/10.1016/s0165-0114(97)00377-1, 2000.

Chiew, F.H.S., Zhou, S.L. and McMahon, T.A.: Use of seasonal streamflow forecasts in water resources management, *Journal of Hydrology*, 270(1-2), 135–144, https://doi.org/10.1016/s0022-1694(02)00292-5, 2003.

Chiu, W.-Y., Yen, G.G. and Juan, T.-K.: Minimum Manhattan distance approach to multiple criteria decision making in multiobjective optimization problems, *IEEE Transactions on Evolutionary Computation*, 20(6), 972–985, https://doi.org/10.1109/tevc.2016.2564158, 2016.

Copernicus: Climate Data Store, https://cds.climate.copernicus.eu/ (last access: 19 May 2024), 2024.

Crippa, N., Grillakis, M.G., Tsilimigkras, A., Yang, G., Giuliani, M. and Koutroulis, A.G.: Seasonal forecast-informed reservoir operation, Potential benefits for a water-stressed Mediterranean basin, *Climate Services*, 32, 100406–100406, https://doi.org/10.1016/j.cliser.2023.100406, 2023.

701 Crochemore, L., Ramos, M.-H. and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill
702 of seasonal streamflow forecasts, *Hydrology and Earth System Sciences*, 20(9), 3601–3618,
703 https://doi.org/10.5194/hess-20-3601-2016, 2016.

704 Das, I.: On characterizing the 'knee' of the Pareto curve based on normal-boundary intersection, *Structural
705 Optimization*, 18(2), 107–115, 1999.

706 Day, G.N.: Extended streamflow forecasting using NWSRFS, *Journal of Water Resources Planning and
707 Management*, 111(2), 157–170, https://doi.org/10.1061/(asce)0733-9496(1985)111:2(157), 1985.

708 Ehsani, N., Vörösmarty, C.J., Fekete, B.M. and Stakhiv, E.Z.: Reservoir operations under climate change:
709 Storage capacity options to mitigate risk, *Journal of Hydrology*, [online] 555, 435–446,
710 https://doi.org/10.1016/j.jhydrol.2017.09.008, 2017.

711 Fan, F.M., Schwanenberg, D., Alvarado, R., Reis, A.A., Collischonn. W. and Naumman. S.: Performance of
712 deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower
713 reservoir, *Water Resources Management*, 30, 3609–3625, https://doi.org/10.1007/s11269-016-1377-8, 2016.

714 Ficchì, A., Raso, L., D Dorchies, Pianosi, F., Malaterre, P-O., van Overloop, P-J. and Jay-Allemand, M.:
715 Optimal operation of the multireservoir system in the Seine River basin using deterministic and ensemble
716 forecasts, *Journal of Water Resources Planning and Management*, 142(1),
717 https://doi.org/10.1061/(asce)wr.1943-5452.0000571, 2016.

718 Fishburn, P.C.: Additive utilities with finite sets: Applications in the management sciences, *Naval Research
719 Logistics Quarterly*, 14(1), 1–13, https://doi.org/10.1002/nav.3800140102, 1967.

720 Giagkiozis, I. and Fleming, P.J.: Pareto front estimation for decision making, *Evolutionary Computation*, 22(4),
721 651–678, https://doi.org/10.1162/evco_a_00128, 2014.

722 Goldsmith, E. and Hildyard, N.: *The Social and Environmental Effects of Large Dams*, Random House (NY),
723 1984.

724 Goodarzi, M., Jabbarian Amiri, B., Azarneyvand, H., Khazaee, M. and Mahdianzadeh, N.: Assessing the
725 performance of a hydrological Tank model at various spatial scales, *Journal of Water Management Modeling*,
726 **29**, 1–8, https://doi.org/10.14796/jwmm.c472, 2020.

727 Greuell, W., Franssen, W.H.P., Biemans, H. and Hutjes, R.W.A.: Seasonal streamflow forecasts for Europe –
728 Part I: Hindcast verification with pseudo- and real observations, Hydrology and Earth System Sciences, 22(6),
729 3453–3472, https://doi.org/10.5194/hess-22-3453-2018, 2018.

730 Hurkmans, R.T.W.L., Hurk, B., Schmeits, M., Wetterhall, F. and Pechlivanidis, I.G.: Seasonal streamflow
731 forecasting for fresh water reservoir management in the Netherlands: An assessment of multiple prediction
732 systems, *Journal of Hydrometeorology*, 24(7), 1275–1290, https://doi.org/10.1175/jhm-d-22-0107.1, 2023.

733 Hwang, C.L. and Yoon, K.: *Multiple attribute decision making: methods and applications, A state-of-the-art
734 survey*, Springer-Verlag, New York, 1981.

735 Jackson-Blake, L., Clayer, F., Haande, S., James E.S. and Moe, S.J.: Seasonal forecasting of lake water quality
736 and algal bloom risk using a continuous gaussian bayesian network, *Hydrology and Earth System Sciences*,
737 26(12), 3103–3124, https://doi.org/10.5194/hess-26-3103-2022, 2022.

738 Johnson, F. and Sharma, A.: A nesting model for bias correction of variability at multiple time scales in general
739 circulation model precipitation simulations, *Water Resources Research*, 48(1), 1–16,
740 https://doi.org/10.1029/2011wr010464, 2012.

741 Jung, Y., Nam, W.S., Shin, H. and Heo, J.-H.: A study on low-flow frequency analysis using dam inflow,
742 *Journal of The Korean Society of Civil Engineers*, 32(6B), 363–371,
743 https://doi.org/10.12652/ksce.2012.32.6b.363, 2012.

744 K-water (Korea Water Resources Corporation): *My water*, [online] www.water.or.kr. Available at:
745 http://water.or.kr [Accessed 10 May. 2023], 2023.

746 K-water (Korea Water Resources Corporation): *2013-2018 Sustained drought analysis and assessment report*,
747 South Korea, 2018.

748 Lee, Y., Peñuela, A., Pianosi, F. and Rico-Ramirez, M.A.: Catchment-scale skill assessment of seasonal
749 precipitation forecasts across South Korea, *International journal of climatology*, 43(11), 5092–5111,
750 https://doi.org/10.1002/joc.8134, 2023.

Lee, Y., Pianosi, F., Peñuela, A. and Rico-Ramirez, M. A.: Skill of seasonal flow forecasts at catchment-scale: an assessment across South Korea, *Hydrology and Earth System Science*, 28, 3261–3279, https://doi.org/10.5194/hess-28-3261-2024, 2024.

Li, W., Zhang, G., Zhang, T. and Huang, S.: Knee point-guided multiobjective optimization algorithm for microgrid dynamic energy management, *Complexity*, 2020, 1–11, https://doi.org/10.1155/2020/8877008, 2020.

Liu, P.: Multi-attribute decision-making method research based on interval vague set and TOPSIS method, *Technological and Economic Development of Economy*, 15(3), 453–463, https://doi.org/10.3846/1392-8619.2009.15.453-463, 2009.

Lu, L. anderson-Cook, C.M. and Robinson, T.J.: Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier, *Technometrics*, 53(4), 353–365, https://doi.org/10.1198/tech.2011.10087, 2011.

Lucatero, D., Madsen, H., Refsgaard, J.C., Kidmose, J. and Jensen, K.H.: Seasonal streamflow forecasts in the Ahlergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical consistency, *Hydrology and Earth System Sciences*, 22(7), 3601–3617, https://doi.org/10.5194/hess-22-3601-2018, 2018.

Malekmohammadi, B., Zahraie, B. and Kerachian, R.: Ranking solutions of multi-objective reservoir operation optimization models using multi-criteria decision analysis, *Expert Systems with Applications*, 38(6), 7851–7863, https://doi.org/10.1016/j.eswa.2010.12.119, 2011.

Matheson, J.E. and Winkler, R.L.: Scoring rules for continuous probability distributions, *Management Science*, 22(10), 1087–1096, https://doi.org/10.1287/mnsc.22.10.1087, 1976.

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M. and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, 48(3), 1–34, https://doi.org/10.1029/2009rg000314, 2010.

Millner, A. and Washington, R.: What determines perceived value of seasonal climate forecasts? A theoretical analysis, *Global Environmental Change*, 21(1), 209–218, https://doi.org/10.1016/j.gloenvcha.2010.08.001, 2011.

Mishra, A.K. and Singh, V.P.: A review of drought concepts, *Journal of Hydrology*, [online] 391(1-2), 202–216, https://doi.org/10.1016/j.jhydrol.2010.07.012, 2010.

Ni, X., Dong, Z., Jiang, Y., Xie, W., Yao, H. and Chen, M.: A subjective-objective integrated multi-objective decision-making method for reservoir operation featuring trade-offs among non-inferior solutions themselves, *Journal of Hydrology*, 613, 128430, https://doi.org/10.1016/j.jhydrol.2022.128430, 2022.

Ou, X., Gharabaghi, B., McBean, E. and Doherty, C.: Investigation of the Tank model for urban storm water management, *Journal of Water Management Modeling*, 25, 1–5, https://doi.org/10.14796/jwmm.c421, 2017.

Park, J.Y. and Kim, S.J.: Potential impacts of climate change on the reliability of water and hydropower supply from a multipurpose dam in South Korea, *Journal of the American Water Resources Association*, 50(5), 1273–1288, https://doi.org/10.1111/jawr.12190, 2014.

Park, M., Lee, J.H., Lim, Y.K. and Kwon, H.H.: Estimation of evaporation from water surface in Yongdam dam using the empirical evaporation equation, *Journal of Korea Water Resources Association*, 2, 139–150, 2024.

Pechlivanidis, I.G., Crochemore, L., Rosberg, J. and Bosshard, T.: What are the key drivers controlling the quality of seasonal streamflow forecasts?, *Water Resources Research*, 56(6), https://doi.org/10.1029/2019wr026987, 2020.

Peñuela, A., Hutton, C. and Pianosi, F.: Assessing the value of seasonal hydrological forecasts for improving water resource management: insights from a pilot application in the UK, *Hydrology and Earth System Sciences*, 24(12), 6059–6073, https://doi.org/10.5194/hess-24-6059-2020, 2020a.

Peñuela, A. and Pianosi, F.: iRONS (interactive Reservoir Operation Notebooks and Software), Zenodo, https://doi.org/10.5281/zenodo.4277646, 2020b.

Peñuela, A., Hutton, C. and Pianosi, F.: An open-source package with interactive Jupyter Notebooks to enhance the accessibility of reservoir operations simulation and optimization, *Environmental Modelling & Software*, 145, 105188, https://doi.org/10.1016/j.envsoft.2021.105188, 2021.

Prudhomme, C., Hannaford, J., Harrigan, S., Boorman, D., Knight, J., Bell, V., Jackson, C., Svensson, C., Parry, S., Bachiller-Jareno, N., Davies, H., Davis, R., Mackay, J., McKenzie, A., Rudd, A., Smith, K., Bloomfield, J., Ward, R. and Jenkins, A.: Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales, *Hydrological Sciences Journal*, 62(16), 2753–2768, https://doi.org/10.1080/02626667.2017.1395032, 2017.

Rougé, C., Penuela-Fernandez, A., Pianosi, F.: Forecast families: A new method to systematically evaluate the benefits of improving the skill of an existing forecast, *Journal of Water Resources Planning and Management*, 149(5), https://doi.org/10.1061/JWRMD5.WRENG-5934, 2023.

Ryoo, K.-S., Lee, H.-G., Park, J.-H. and Hur, Y.-T.: Improvement of estimation method on the low flow frequency inflow for the optimal reservoir operation, *Journal of Water Resources Planning and Management*, 1287–1291, 2009.

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

Schwalm, C.R. anderegg, W.R.L., Michalak, A.M., Fisher, J.B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J.D., Wolf, A., Huntzinger, D.N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A. and Tian, H.: Global patterns of drought recovery, *Nature*, 548(7666), 202–205, https://doi.org/10.1038/nature23021, 2017.

Shah, D., Zhao, G., Li, Y., Singh, V.P. and Gao, H.: Assessing global reservoir-based hydrological droughts by fusing storage and evaporation, *Geophysical Research Letters*, 51(1), 1–11, https://doi.org/10.1029/2023gl106159, 2024.

Sheffield, J., Wood, E.F. and Roderick, M.L.: Little change in global drought over the past 60 years, *Nature*, [online] 491(7424), 435–438, https://doi.org/10.1038/nature11575, 2012.

Shiau, J.-T.: Risk-aversion optimal hedging scenarios during droughts, *Applied Water Science*, 13(1), 1–14, https://doi.org/10.1007/s13201-022-01817-x, 2022.

Shrestha, M., Acharya, S.C. and Shrestha, P.K.: Bias correction of climate models for hydrological modelling - are simple methods still useful?, *Meteorological Applications*, 24(3), 531–539, https://doi.org/10.1002/met.1655, 2017.

Soares, B.M. and Dessai, S.: Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe, *Climatic Change*, 137(1-2), 89–103, https://doi.org/10.1007/s10584-016-1671-8, 2016.

Sugawara, M., Watanabe, I., Ozaki, E. and Katsuyama, Y.: *Tank model programs for personal computer and the way to use*, National Research Centre for Disaster Prevention, Japan, 1986.

Sugawara, M.: *"Tank model." Computer models of watershed hydrology*, V. P. Singh (Ed.), Water Resources Publications, Highlands Ranch, Colorado, 1995.

Tian, F., Li, Y., Zhao, T., Hu, H., Pappenberger, F., Jiang, Y. and Lu, H.: Evaluation of the ECMWF system 4 climate forecasts for streamflow forecasting in the Upper Hanjiang River basin, *Hydrology Research*, 49(6), 1864–1879, https://doi.org/10.2166/nh.2018.176, 2018.

Turner, S.W.D., Bennett, J.C., Robertson, D.E. and Galelli, S.: Complex relationship between seasonal streamflow forecast skill and value in reservoir operations, *Hydrology and Earth System Sciences*, 21(9), 4841–4859, https://doi.org/10.5194/hess-21-4841-2017, 2017.

Tzeng, G.H and Huang, J.: *Multiple attribute decision making: methods and applications*, Boca Raton (Florida): Crc Press, 2011.

University of Bristol: SEAFORM, Zenodo [code], https://doi.org/10.5281/zenodo.12800811, 2023a.

University of Bristol: SEAFLOW, Zenodo [code], https://doi.org/10.5281/zenodo.12800917, 2023b.

Vassoney, E., Mammoliti Mochet, A., Erika, D., Negro, G., Pilloni, M.G. and Comoglio, C.: Comparing multi-criteria decision-making methods for the assessment of flow release scenarios from small hydropower plants in the alpine area, *Frontiers in Environmental Science*, 9, 1–20, https://doi.org/10.3389/fenvs.2021.635100, 2021.

Wang, Z. and Rangaiah, G.P.: Application and analysis of methods for selecting an optimal solution from the Pareto-optimal front obtained by multiobjective optimization, *Industrial & Engineering Chemistry Research*, 56(2), 560–574, https://doi.org/10.1021/acs.iecr.6b03453, 2017.

20

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

854 Wurbs, R.A. and Ayala, R.A.: Reservoir evaporation in Texas, USA, *Journal of Hydrology*, [online] 510, 1–9,
855 https://doi.org/10.1016/j.jhydrol.2013.12.011, 2014.

856 Yang, G., Guo, S., Liu, P. and Block, P.: Sensitivity of forecast value in multiobjective reservoir operation to
857 forecast lead time and reservoir characteristics, *Journal of Water Resources Planning and Management*, 147(6),
858 https://doi.org/10.1061/(asce)wr.1943-5452.0001384, 2021.

859 Yoe, C.E.: *Principles of risk analysis: decision making under uncertainty*, Boca Raton, Fl: Crc Press, Taylor
860 And Francis, 2019.

861 You, G.J.-Y.: Hedging Rules for the operation of lake Okeechobee in Southern Florida, *Journal of the American*
862 *Water Resources Association*, 49(5), 1179–1197, https://doi.org/10.1111/jawr.12078, 2013.

863 Yossef, N.C., Winsemius, H., Weerts, A., van Beek, R. and Bierkens, M.F.P.: Skill of a global seasonal
864 streamflow forecasting system, relative roles of initial conditions and meteorological forcing, *Water Resources*
865 *Research*, 49, 4687–4699, https://doi.org/10.1002/wrcr.20350, 2013.

866 Zhang, X., Hao, Z., Singh, V.P., Zhang, Y., Feng, S., Xu, Y. and Hao, F.: Drought propagation under global
867 warming: Characteristics, approaches, processes and controlling factors, *Science of The Total Environment*,
868 [online] 838, 156021, https://doi.org/10.1016/j.scitotenv.2022.156021, 2022.

869 Zhao, T., Cai, X. and Yang, D.: Effect of streamflow forecast uncertainty on real-time reservoir operation,
870 *Advances in Water Resources*, 34(4), 495–504, https://doi.org/10.1016/j.advwatres.2011.01.004, 2011.

871 Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P.N. and Zhang, Q.: Multiobjective evolutionary
872 algorithms: A survey of the state of the art, *Swarm and Evolutionary Computation*, 1(1), 32–49,
873 https://doi.org/10.1016/j.swevo.2011.03.001, 2011.

874 Zhu, F., Zhong, P., Sun, Y. and Xu, B.: Selection of criteria for multi-criteria decision making of reservoir flood
875 control operation, *Journal of Hydroinformatics*, 19, 558–571, https://doi.org/10.2166/hydro.2017.059, 2017.

876 Zhu, F., Zhong, P., Xu, B., Wu, Y. and Zhang, Y.: A multi-criteria decision-making model dealing with
877 correlation among criteria for reservoir flood control operation, *Journal of Hydroinformatics*, 18(3), 531–543,
878 https://doi.org/10.2166/hydro.2015.055, 2015.