https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020

Hydrology and Earth System Sciences
Discussions

Assessing the value of seasonal hydrological forecasts for improving water resource management: insights from a pilot application in the UK

Andres Peñuela[1], Christopher Hutton[2], Francesca Pianosi[1, 3]

[1]Civil Engineering, University of Bristol, Bristol, BS8 1TR, UK
[2]Wessex Water Services Ltd, Bath, BA2 7WW, UK
[3]Cabot Institute, University of Bristol, BS8 1UH, UK

Correspondence to: Andres Peñuela (andres.penuela-fernandez@bristol.ac.uk)

Abstract. Improved skill of long-range weather forecasts has motivated an increasing effort towards developing seasonal hydrological forecasting systems across Europe. Among other purposes, such forecasting systems are expected to support better water management decisions. In this paper we evaluate the potential use of a real-time optimisation system (RTOS) informed by seasonal forecasts in a water supply system in the UK. For this purpose, we simulate the performances of the RTOS fed by ECMWF seasonal forecasting systems (SEAS5) over the past ten years, and we compare them to a benchmark operation that mimics the common practices for reservoir operation in the UK. We also attempt to link the improvement of system performances, i.e. the forecast value, to the forecast skill (measured by the mean error and the Continuous Ranked Probability Skill Score) as well as other factors such as bias correction, the decision maker priorities, hydrological conditions and level of uncertainty consideration. We find that some of these factors control the forecast value much more strongly than the forecast skill. For the (realistic) scenario where the decision-maker prioritises water resource availability over energy cost reductions, we identify clear operational benefits from using seasonal forecasts, provided that forecast uncertainty is explicitly considered. However, when comparing the use of ECMWF-SEAS5 products to ensemble streamflow predictions (ESP), which are more easily derived from historical weather data, we find that ESP remains a hard-to-beat reference not only in terms of skill but also in terms of value.

1. Introduction

In a water-stressed world, where water demand and climate variability are increasing, it is essential to improve the efficiency and lifespan of existing water infrastructure along with, or possibly in place of, developing new one (Gleick, 2003). In the current information age, there is a great opportunity to do this by improving the ways in which we use hydrological data and simulation models (the 'information infrastructure') to inform operational decisions (Gleick et al., 2013, Boucher et al., 2012).

Hydro-meteorological forecasting systems are a prominent example of information infrastructure that has a huge potential for improving water infrastructure operation efficiency. The usefulness of hydrological forecasts has been demonstrated in several applications, particularly to enhance reservoir operations for flood management (Voisin et al., 2011, Wang et al.,

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

2012, Ficchì et al., 2016) and hydropower production (Faber and Stedinger, 2001, Maurer and Lettenmaier, 2004, Alemu et al., 2010, Fan et al., 2016). In these types of systems, we usually find a strong relationship between the forecast skill (i.e. the forecast ability to anticipate future hydrological conditions) and the forecast value (i.e. the improvement in system performance obtained by using forecasts to inform operational decisions). However, this relationship becomes weaker for water supply systems, in which the storage buffering effect may reduce the importance of the forecast skill (Anghileri et al., 2016, Turner et al., 2017), particularly when the reservoir capacity is large (Maurer and Lettenmaier, 2004, Turner et al., 2017). Moreover, in water supply systems, decisions are made taking into consideration the hydrological conditions over lead time of several weeks or even months. Forecast products with such lead times, i.e. 'seasonal' forecasts, are typically less skilful compared to the short or medium range forecasts used for flood control or hydropower production applications.

When using seasonal hydrological scenarios or forecasts to assist water system operations, three main approaches are available: worst case scenario, ensemble streamflow prediction (ESP) and dynamical streamflow prediction (DSP). In the worst-case scenario approach, operational decisions are made by simulating their effects against a repeat of the worst hydrological droughts on records. Worst-case forecasts clearly have no particular skill, but their use has the advantage that it provides a lower bound of system performance and reflect the risk-adverse attitude of most water resource management practice. This approach is commonly applied by water companies in the UK for reservoir operation and it is recommended by the water resource management guidelines of the UK Environment Agency (EA, 2017).

In the ensemble streamflow prediction (ESP) approach, a hydrological forecasts ensemble is produced by forcing a hydrological model using the current initial hydrological conditions and historical weather data over the period of interest (Day, 1985). Operational decisions are then evaluated against such ensemble. The skill of the ESP ensemble is mainly due to the updating of the initial conditions. However, since ESP is limited to the range of past observations, ESP forecasts can have limited skill under non-stationary climate and where initial conditions do not dominate the seasonal hydrological response (Arnal et al., 2018). Nevertheless, the ESP approach is popular among operational agencies thanks to its simplicity, low cost, efficiency and its intuitively appealing nature (Bazile et al., 2017), i.e. ESP is coherent to the human tendency to examine a situation according to past experiences. Seasonal ESP was used to assess possible improvements of supply-hydropower systems operation, e.g. by Alemu et al. (2010) who reported achieving an average economic benefit of 7% with respect to the benchmark operation policy, and by Anghileri et al. (2016) who however did not observe significant improvements (possibly because they only used the ESP mean, instead of the full ensemble).

Last, the dynamical streamflow prediction (DSP) approach uses seasonal weather forecasts produced by a dynamic climate model to feed the hydrological model (instead of historical weather data). The output is also an ensemble hydrological forecast, whose skill comes from the updated initial condition as well as the predictive ability of numerical weather forecasts, due to global climate teleconnections such as the El Niño Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). Therefore, these forecasts are generally more skilful in areas where climate teleconnections exert a strong influence, such as tropical areas, and particularly in the first month ahead (Block and Rajagopalan, 2007). In areas where climate teleconnections have a weak influence, instead, DSP can have lower skill than ESP, particularly beyond the first lead month

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

(Arnal et al., 2018, Greuell et al., 2019). Nevertheless, recent advances in the prediction of climate teleconnections in Europe, such as the NAO (Wang et al., 2017, Scaife et al., 2014, Svensson et al., 2015) means that seasonal forecasts skill is likely to continue increasing in next years. Post-processing techniques such as bias correction can also potentially improve seasonal streamflow forecast skill (Crochemore et al., 2016). However, studies assessing the benefits of bias correction for seasonal hydrological forecasting are still rare in the literature, while studies on long-term hydrological projections (Ehret et al., 2012, Hagemann et al., 2011) highlighted a lack of clarity on whether bias correction should be applied or not. In recent years, meteorological centres such as the European Centre for Medium-Range Weather Forecast (ECMWF) and the UK Met Office, have made important efforts to provide skilful seasonal forecasts, both meteorological (Hemri et al., 2014, MacLachlan et al., 2015) and hydrological (Bell et al., 2017, Arnal et al., 2018) in the UK and Europe, and encouraged their application for water resource management. To our knowledge, however, pilot applications demonstrating the value of such seasonal forecast products to improve operational decisions are still lacking.

While the skill of DSP is likely to keep increasing in the next years, this may still not produce considerable improvement in water system operations soon, especially in water supply systems where the forecast skill-value relationship is weaker. Nevertheless, a number of studies have demonstrated that other factors, which are not necessarily captured by forecast skill scores, may also be important to improve the value of short-term and seasonal forecasts. These include accounting for forecast uncertainty in the system operation optimization (Yao and Georgakakos, 2001, Boucher et al., 2012, Fan et al., 2016), using less rigid operation approaches (Yao and Georgakakos, 2001, Brown et al., 2015, Georgakakos and Graham, 2008) and making optimal operational decisions during severe droughts (Turner et al., 2017). Additionally, the forecast skill itself can be defined in different ways, and it is likely that different characteristics of forecast errors (sign, amount, timing, etc.) affect the forecast value in different ways. Widely used skill scores for hydrological forecast ensembles are the rank histogram (Anderson, 1996), the relative operating characteristic (Mason, 1982) and the ranked probability score (Epstein, 1969). The ranked probability score is widely used by meteorological agencies and it is the recommended score for evaluation of overall performance since it provides a measure of both the bias and the spread of the ensemble into a single factor, while it can also be decomposed into different sub-factors in order to look at the different attributes of the ensemble forecast (Pappenberger et al., 2015, Arnal et al., 2018). However, whether these skill score definitions are relevant for the specific purpose of water resources management, or other definitions would be better proxy of the forecast value, remains an open question.

In this paper, we aim at assessing the value of DSP for improving water system operation by application to a real-world reservoir system, and in doing so we build on this growing effort to improve seasonal hydrometeorological forecasting systems and make them suitable for operational use in the UK (Bell et al., 2017, Prudhomme et al., 2017). Through this application we aim to answer the three following questions: 1) can the efficiency of a UK real-world reservoir supply system be improved by using DSP?, 2) does accounting for forecast uncertainty improve forecast value (for the same skill)? and 3) what other factors influence the forecast skill-value relationship?

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

For this purpose, we will simulate and compare the performance of a real-time optimization system informed by seasonal weather forecasts over a historical period for which both observational and forecast datasets are available, and we will benchmark it to a worst-case scenario approach, which is commonly used to inform water supply management in the UK. As for the seasonal forecast products, we will assess both ESP and DSP derived from the ECMWF seasonal forecast products (Tim et al., 2018). We will also compare the forecast skill and value before and after applying bias correction to the ECMWF forecast products, and for different degrees of forecast uncertainty (i.e. different ensemble sizes). To account for decision-making uncertainty, we will also simulate the performance of the system under five operating scenarios representing different operational priorities. Finally, we will discuss opportunities and barriers to bring such approach into practice.

Our results are meant to provide water managers an evaluation of the potential of using seasonal forecasts in extra-tropical areas, such as the UK, and to give forecasts providers indications on directions for future developments that may make their products more valuable for water management.

2. Methodology

2.1. Real-time optimization system

An overview of the real-time optimization system (RTOS) informed by seasonal weather forecasts is given in Figure 1 (left part). It consists of three main stages that are repeated each time an operational decision must be made. These three stages are:

1.a Forecast generation. We use a hydrological model forced by seasonal weather forecasts to generate the seasonal hydrological forecasts. The initial conditions are determined by forcing the same model by (recent) historical weather data for a warm-up period. Another model determines the future water demand during the forecast horizon. Although not tested in this study, in principle such demand model could also be forced by seasonal weather forecast.

1.b Optimization. This stage uses (i) a reservoir system model to simulate the reservoir storages in response to given inflows and operational decisions, (ii) a set of operation objective functions to evaluate the performance of the system, and (iii) an optimizer to determine the set of optimal operational decisions that realise optimal trade-offs between the objective functions.

1.c Selection of one trade-off solution. In this stage, we represent the performance of the optimal trade-off decisions in what we call a “pre-evaluation Pareto front”. The terms “pre-evaluation” highlights that these are the anticipated performances according to our models and hydrometeorological forecasts, not the actual performances achieved when the decisions are implemented (which are unknown at this stage). Among this set of optimal decisions, the operator will select one according to their priorities, i.e. the relative importance given to each operation objectives. In a simulation experiment, we can mimic the operator choice by setting some rule to choose one point on the Pareto front (and apply it consistently at each decision timestep of the simulation period).

2.2. Evaluation

When the RTOS is implemented in practice, the selected operational decision is applied to the real system and the RTOS used again, with updated system conditions, when a new decision needs to be made or new weather forecasts become

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

available. If however we want to evaluate the performance of RTOS in a simulation experiment (for instance to demonstrate the value of using RTOS to reservoir operators) we need to combine it with the evaluation system depicted in the right part of Figure 1. Here, the selected operational decision coming out of the RTOS is applied to the reservoir system model, instead of the real system. The reservoir model is now forced by hydrological inputs observed in the (historical) simulation period, instead of the seasonal forecasts, which enables us to estimate the actual flows and next-step storage that would have occurred if the RTOS was used at the time. This simulated next-step storage can then be used as the initial storage volume for running the RTOS at the following timestep. Once the process has been repeated for the entire period of study, we can provide an overall evaluation of the hydrological forecast skill and the performance of the RTOS, i.e. the forecast value. This evaluation (Figure 1) consists of two stages:

2.a Forecast skill evaluation. In order to evaluate the capacity of the hydrological forecast to predict the observed inflows we apply forecast skill scores and absolute error indicators. In this paper, we will use the continuous ranked probability skill score (CRPSS) and the absolute difference between the observed and forecasted inflows.

2.b Forecast value evaluation. The forecast value is presented as the improvement of the system performance obtained by using the RTOS over the simulation period, with respect to the performance under a benchmark operation. Notice that, because the RTOS deals with a multi-objective problem and we have to implement a rule to select one solution out of the pre-evaluation Pareto front, in principle we could run a different simulation experiment for each possible definition of the selection rule, i.e. for each possible definition of the operational priorities. However, for the sake of simplicity, we only simulate five different operational priorities, and thus obtain a post-evaluation Pareto front with five points. In this Pareto front the origin of the coordinates represents the performance of the benchmark operation. Therefore, a positive value along one axis represents an improvement in that operation objective with respect to the benchmark, whereas a negative value represents a deterioration. When values are positive on both axes, the simulated RTOS solution dominates (in a Pareto sense) the benchmark; the further away from the origin, the more the forecast has proven valuable for decision-making. If instead one value is positive and the other is negative then we would conclude that the forecast value is neither positive nor negative, because the improvement of one objective by the RTOS was achieved at the expenses of the other.

2.3. Case study

2.3.1. Description of the reservoir system

The reservoir system used in this case study is a two-reservoir system in the South West of the UK (schematised in Figure 2). The two reservoirs are moderately sized with storage capacities in the order of 20,000 megalitres (Ml) (S1) and 5,000 Ml (S2) (the average of UK reservoirs is 1,377 Ml (EA, 2017)). The system is partially shared between two different water companies, reservoir S1 being the system element used by both companies. The gravity releases from this reservoir ($u_{S1,R}$) are used by the owner company to support downstream abstraction during low river (R) flows. The other company can also use pumped releases from S1 ($u_{S1,D}$) to complement gravity releases ($u_{S2,D}$) from their own reservoir (S2) in supplying D in a wider conjunctive use system. Both reservoirs are required to make environmental compensation releases.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

A key operational aspect of the system is the possibility of pumping water into the shared reservoir S1. Pumped inflows ($u_{R,S1}$) may be operated in the winter months to supplement natural inflows, provided sufficient water is available in the river. This facility provides additional drought resilience, as it allows the companies to increase reservoir storage if natural inflows are insufficient during the winter months (from 1st November till 1st April) to ensure meeting the summer demand. The two companies that operate the system liaise regularly, particularly regarding the pumped storage operation, which is constrained by rule curves, and has operated in eleven years since 1995. As the pump energy consumption is costly, there is an important trade-off between the operating cost of pump storage and achieving drought resilience.

The rule curve applied in the current operation procedures defines the storage level at which pumps are triggered. Each point on the curve is derived based on the amount of pumping required to refill the reservoir under the worst historical observed inflows between that point in time and the end of the pump storage period (1st April). The pumping trigger is therefore risk-averse, which means there is a reasonable change of pumping too early on during the refill period. This increases the likelihood of reservoir spills if spring rainfall is abundant, which means unnecessary expenditure on pumping. Informing the pump operations by using seasonal forecasts of future natural inflows (I_{S1} and I_{S2}) may thus help to reduce the volume of water pumped whilst achieving the same reservoir storage at the end of the refilling period.

2.3.2. Forecast generation

In this study we generated dynamical streamflow predictions (DSP) by forcing a lumped hydrological model, the HBV model (Bergström and Singh, 1995), with the seasonal ECMWF SEAS5 weather hindcasts (Tim et al., 2018). The ECMWF SEAS5 dataset consists of an ensemble of 25 members starting on the 1st day of every month and providing daily temperature and precipitation with a lead time of 7 months. The spatial resolution is 36 km which compared to the catchment sizes (28.8 km^2 for S1 and 18.2 km^2 for S2) makes it necessary to bias correct and downscale the ECMWF hindcasts. Given the lack of clarity in the potential benefits of bias correction (Ehret et al., 2012), we will provide results of using both non-corrected and bias corrected forecasts. . The dataset of weather hindcast is available from 1981, whereas reservoir data are available for the period 2005-2016. Hence, we used the period 2005-2016 for the RTOS evaluation and the earlier data from 1981 for bias correction. While limited, this period captures a variety of hydrological conditions, including dry ones in 2005-06, 2010-2011 and 2011-12, relatively close to the driest period on records (1975-1976) (more in (Figure 8) of the Supplementary Material). This is important because under drier conditions, the system performance is more likely to depend on the forecast skill and the benefits of RTOS may become more apparent (Turner et al., 2017).Daily inflows were converted to weekly inflows for consistency with the weekly time step applied in the reservoir system model.

A linear scaling approach (or “monthly mean correction”) was applied for bias correction. This approach is simple and often provides similar results as more sophisticated approaches such as the quantile or distribution mapping (Crochemore et al., 2016). A correction factor is calculated as the ratio between the average daily observed and forecasted (ensemble mean) values of the variable of interest (precipitation or temperature) for a given month and year. The correction factor is then applied as a multiplicative factor to correct the raw daily forecast values. A different factor is calculated and applied for each month and each year of the evaluation period (2005-2016). For example, for November 2005 we obtain the correction factor

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

as the ratio between the mean observed rainfall in November from 1981 until 2004 (i.e. the average of 24 values) and the mean forecasted rainfall for the same months (i.e. the average of 24x25 values, as we have 25 ensemble members). For November 2006, we re-calculate the correction factor by also including the observations and forecasts of November 2005, hence taking averages over 25 values; and so forth.

As anticipated in the Introduction, the ESP is an ensemble of equiprobable weekly streamflow forecasts generated by the hydrological model (HBV in our case) forced by meteorological inputs observed in the past. In our case and for consistency with what done for the bias correction of ECMWF SEAS5 forecasts, we use meteorological observations (precipitation and temperature) from 1981 until the year before the simulated decision timestep to produce the ESP. This also produces an ensemble of similar size (24 to 35 members) with respect to the ECMWF ensemble (25 members).

2.3.3. Optimization: Reservoir system model, operation objective functions and optimiser

The reservoir system dynamics is simulated by a mass balance model implemented in Python. The simulation model is linked to an optimiser to determine the optimal scheduling of pumping ($u_{R,S1}$) and release ($u_{S1,D}$ and $u_{S2,D}$) decisions. As optimiser we used the NSGA-II multi-objective evolutionary algorithm (Deb et al., 2002) implemented in the open-source python package Platypus (Hadka, 2018). We set two operation objectives for the optimiser: to minimize pumping energy costs and to maximize the water resource availability at the end of the pump storage period. The first objective function is calculated as the sum of the energy costs associated to pumped inflows and pumped releases ($u_{R,S1}$ and $u_{S1,D}$) over the optimisation period. The second function is the mean storage volume in S1 and in S2 at the end of the optimisation period (1st April). The release of S1 $u_{S1,R}$ is not considered a decision variable and is defined by the observed values during the period of study. This choice is however not likely to have important implications on the optimization results because $u_{S1,R}$ on average only represents the 15% of the total S1 releases ($u_{S1,D}$ + $u_{S1,R}$). Also, we made the simplifying assumption that the future water demand is perfectly known at each time step, and thus defined D by the sum of the observed releases from S1 ($u_{S1,D}$) and S2 ($u_{S2,D}$) for the period of study, instead of using a demand model. The simplification enables us to focus on the relationship between seasonal hydrological forecast skill and the forecast value while avoiding the influence of non-perfect water demand forecasts. More details about the reservoir simulation model are given in the Supplementary Material.

2.3.4. Selection of trade-off solution

In order to take into account the uncertainty in the selection of the trade-off solution, we estimate the forecast value under five operating scenarios out of the 20 available in the pre-evaluation Pareto front (see Figure 1). They represent five selection rules based on different operational priorities, according to the relative importance given to each performance objectives: 1) resource availability only (*rao*), 2) resource availability prioritised (*rap*), 3) balanced (*bal*), 4) pumping savings prioritised (*psp*) and 5) pumping savings only (*pso*). The same selection rule is consistently applied at each decision timestep of the simulation period. The relative importance of the objectives is quantified as the percentile of the performance improvement along the axes of the pre-evaluation Pareto front. For instance, *rao* is the extreme solution in the pre-evaluation Pareto front that delivers the largest improvement in resource availability; *rap* is the solution delivering the 75% percentile in resource availability increase among the 20 operation scenarios available; *bal* delivers the median improvement; etc.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

2.3.5. Forecast skill evaluation

We used two metrics to evaluate the forecast skill: a skill score and the mean error.

A skill score evaluates the performance of a given forecasting system with respect to the performance of a reference forecasting system. As a measure of performance, we use the continuous ranked probability score (CRPS) (Brown, 1974) (Hersbach, 2000). As a measure of performance, we use the continuous ranked probability score (CRPS) (Brown, 1974). The skill score is then defined as:

$$CRPSS = 1 - \frac{CRPS^{Sys}}{CRPS^{Ref}}$$

When the skill score is higher (lower) than zero, the forecasting system is more (less) skilful than the reference. When it is equal to zero, the system and the reference have equivalent skill. Following the recommendation by Harrigan et al. (2018) we used ensemble streamflow predictions (ESP) as a "tough to beat" reference, which is more likely to demonstrate the "real skill" of the hydrological forecasting system (Pappenberger et al., 2015).

The continuous ranked probability score appearing in the above equation is defined as the distance between the cumulative distribution function of the probabilistic forecast and the empirical distribution of the corresponding observation. At each forecasting step, and for a given lead time, CRPS is thus calculated as:

$$CRPS(p(x), y) = \int (p(x) - H(x < y))^2 dx$$

Where *p(x)* represents the distribution of the forecast; y is the observation; and *H* is the empirical distribution of the observation, i.e. the step function which equals 0 when $x < y$ and 1 when $x > y$. The lower the CRPS, the better the performance of the forecast. The average CRPS for a given lead time is equal to the mean of the CRPS values across the time frame. In this study weekly forecast and observation data were used to compute CRPS.

The mean error measures the difference between the forecasted and the observed inflows. The mean error is negative when the forecasts tend to underestimate the observations and positive when the forecasts overestimate the observations. The mean error for a given forecasting step and lead time *T* [weeks] is:

$$mean\ error = \frac{1}{M}\sum_{m=0}^{M}\left(\frac{1}{T}\sum_{t=0}^{T}\left(I_{t,m}^{Syst} - I_{t}^{Obs}\right)\right)$$

where *I* is the inflow [ML], *t* is the timestep [week] and *M* the total number of members (*m*) of the ensemble.

2.3.6. Forecast value evaluation and definition of the benchmark operation

To evaluate the forecast value of DSP (before and after bias correction) and ESP, we compared the performance of the RTOS (Figure 1) informed by these seasonal weather forecast products with a benchmark. The benchmark mimics common practices in reservoir operation in the UK, whereby operational decisions are made against a worst-case scenario – a repeat of the worst hydrological drought on records. We can simulate the benchmark operation using similar steps as in the RTOS represented in Figure 1, but with three main variations. First, instead of seasonal weather forecasts, we use the historical weather data recorded in Nov 1975-Apr 1976. Second, the optimiser only determines the optimal scheduling of reservoir releases ($u_{S1,D}$ and $u_{S2,D}$), whereas pumped inflows ($u_{R,S1}$) are determined by the rule curve applied in the current operation

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

procedures. Third, the optimiser only aims at minimising pumping costs, whereas the resource availability objective is turned into a constraint, i.e. the mean storage volume of the two reservoirs must be maximum by the end of the pump storage period (1st April) and no trading-off with pumping costs reduction is allowed.

3. Results

3.1 Forecast skills

First, we analyse the skill of DSP hydrological forecasts. Figure 3a shows the average CRPSS at different lead times before (red) and after (blue) bias correction. Before bias correction, the average forecast skill is highest at 1 month lead time and decreases with larger lead time (solid red line). Furthermore, the skill is higher than average in the three driest winters, i.e. 2005-2006, 2010-2011, 2011-2012 (dashed lines). If we compare DSP to DSP-corr (red and blue solid lines), we see that bias correction deteriorates the average skills for shorter lead times (1 and 2 months) while it improves it for longer ones (3,4 and 5 months). In the driest years (dashed lines) bias correction deteriorates the skill for most lead times.

The average mean error (Figure 3b) indicates that DSP systematically underestimates the inflow observations but less so in the three driest winters. After bias correction (DSP-corr), this systematic underestimation turns into a systematic overestimation. Also, the average mean error gets lower for longer lead times, though not as much in the driest years.

In summary, we can conclude that bias correction does not seem to produce a systematic improvement in the forecast skill for our observation period, but only some improvement at some lead times. On the other hand, what we find in our case study is a clear signal of bias correction turning negative mean errors (inflow underestimation) into positive errors (overestimation). So, while the magnitude of errors stays relatively similar, the sign of those errors changes. We will go back to this point later on, when analysing the skill-value relationship.

3.2 Forecast value

The forecast value is presented here as the simulated system performance improvement, i.e. increase in resource availability and in pumping cost savings, with respect to the benchmark operation.

3.2.1 Effect of operational priority scenario and forecast product on the forecast value

We start by analysing the average forecast value over the simulation period 2005-2016 (Figure 4) for the three seasonal weather forecast products (DSP, DSP-corr and ESP) and the perfect forecast, under five operational policy scenarios (rao: resource availability only; rap: resource availability prioritised; bal: balanced; psp: pumping savings prioritised; and pso: pumping savings only).

Firstly, we notice in Figure 4 that the monthly pumping energy cost savings vary widely with the operational priority. The range of variation depends on the forecast type, going from £20,000 to £48,000 for the perfect forecast and from -£77,000 to £48,000 for the three seasonal weather forecasts. For all forecast products, the improvement in resource availability shows lower variability with an improvement of less than +2% (of the mean storage volume in S1 and in S2 at the end of the optimisation period) for *rao*, and a deterioration of -2% for *pso*. While this seems to suggest a lower sensitivity of the

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

300 resource availability objective, variations of few percent points in storage volume may still be important in critically dry years.

As for the forecast value, we find that perfect forecast brings value (i.e. a simultaneous improvement of both objectives) in the two scenarios that prioritize the increase in resource availability (*rao* and *rap*), DSP brings no value in any scenarios, DSP-corr has positive value in the *rap* and *bal* scenario, and ESP in the *bal* only. In other words, real-time optimisation
305 based on seasonal forecasts can outperform the benchmark operation, but whether this happens depends on both the forecast product being used and the operational priority.

An interesting observation in Figure 4 is that the distance in performance between using perfect forecasts and real forecasts (DSP, DSP-corr, ESP) is very small under scenarios that prioritise energy savings (bottom-right quadrant) and much larger under scenarios prioritising resource availability (top quadrants). This indicates a stronger skill-value relationship under the
310 latter scenarios, i.e. improvements in the forecast skill are more likely to produce improvements in the forecast value if resource availability is the priority.

Last, if we compare DSP with DSP-corr we see that the effect of bias correction is mainly a systematic shift to the right along the horizontal axis, i.e. an improvement in energy cost savings at almost equivalent resource availability. Thanks to this shift, in the scenario that prioritises resource availability (*rap*), DSP-corr outperforms ESP. In fact, using DSP-corr is
315 win-win with respect to the benchmark (i.e. the *rap* performance falls in the top-right quadrant in Figure 4) while using ESP is not, as it improves the resource availability at the expenses of pumping energy savings (i.e. producing negative savings).

3.2.2 Effect of uncertainty consideration on forecast value

We now analyse the effect that different characterisations of forecast uncertainty have on the DPS-corr forecast value. We
320 start by the extreme case when uncertainty is not considered at all in the real-time optimisation, i.e. when we take the mean value of the DSP-corr forecast ensemble and use it to drive a deterministic optimisation. The results are reported in Figure 5, which shows that the solution space shrinks to the bottom-right quadrant and, no matter the decision maker priority, the deterministic forecast has no value because energy savings are only achieved at the expenses of reducing the resource availability.

325 We also consider intermediate cases where optimisation explicitly considers the forecast uncertainty, but the size of the forecast ensemble varies between 5 and 25 members (the original ensemble size). For clarity of illustration, we focus on the resource availability prioritised (*rap*) scenario only. We chose this scenario because it seems to best reflect the current preferences of the system managers, whose priority is to maintain the resource availability while reducing pumping costs as a secondary objective. Moreover, the previous analysis (Figure 4) has shown that the optimised *rap* has a larger window of
330 opportunity for improving performance with respect to the benchmark and could potentially improve both operation objectives if the forecast skill was perfect.

For each chosen ensemble size, we randomly choose 10 replicates of that same size from the original ensemble, then we run a simulation experiment using each of these replicates, and finally average their performance. Results are again shown in

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

Figure 6. For a range of 10 to 20 ensemble members, the forecast value remains relatively close to the value obtained by considering the whole ensemble (25 members). However, if only 5 members are considered, the resource availability is definitely lower and cost savings higher, so that the trade-off that is actually achieved is different from the one that was pursued (i.e. to prioritise resource availability). Notice that the extreme case of using 1 member, i.e. the deterministic forecast case (green cross in Figure 6), further exacerbates this effect of 'achieving the wrong trade-off' as resource availability is even lower than in the benchmark.

3.2.3 Year-by-year analysis of the forecast value

We now study more in detail the year-by-year relationship between skill and value, and between hydrological conditions and value. Again, for the sake of simplicity we focus on the most relevant priority scenario of resource availability priority (*rap*). For this scenario, Figure 6 plots the improvement in system performance achieved in every year against different indicators of skill and hydrological conditions (the plots for the other scenarios are reported in the Supplementary Material).

The two top and bottom panels on the left (a,b, f and g) show that the forecast skill, measured by either the CRPSS or the mean error, is in general weakly correlated to the system performances (Spearman coefficient < 0.5 and p-value > 0.05). Similarly, weak correlation was found in the other priority scenarios (see Supplementary material). The other panels (c-e,h-j) show that the Initial storage (on November, 1st), the Total inflows (from November to the end of April), and their sum (called 'Hydrological conditions') are more strongly correlated to the performance. In particular, the correlation is strongest and with highest confidence (Spearman correlation = -0.60, p-value = 0.05) between the Hydrological conditions and the Increase in resource availability (Figure 6e). The correlation between the Initial storage and the Increase of resource availability (Figure 6c) is lower (Spearman correlation = -0.41, p-value = 0.21), although visually we can observe a threshold effect with a sharp increase of the value in the two years with the lowest initial storage (2011-2012 and 2010-2011). This result may have interesting operational implications, as further discussed in the next Section.

Last, in Figure 7 we investigate the distribution of benefits (i.e. increased resource availability, top, and energy cost savings, bottom) along the simulation period. We compare three different forecast products, DSP, DSP-corr and ESP, in the *rap* scenario. First, we observe that two specific year play the most important role in improving the system performance with respect to the benchmark: 2010-11 for pumping cost savings (bottom panel) and 2011-12 for resource availability (top). These years correspond to the driest conditions in the period of study (see inflow and initial condition data in the top panel, and the Supplementary Material for further analysis of the inflow data). When comparing DSP-corr with DSP (blue and grey bars), we observe that they perform similarly in terms of resource availability but DSP-corr performs better for energy savings. This difference was observed already when looking at average performances over the simulation period (Figure 4) and can be related to the change in sign of forecasting errors induced by bias correction (Figure 3b). In fact, without bias correction, reservoir inflows tend to be underestimated, which leads the RTOS to pump more frequently and often unnecessarily (e.g. in 2005-06, 2006-07, 2007-08, etc.). With bias correction, instead, inflows tend to be overestimated, and the RTOS uses pumping less frequently. Interestingly, the reduction in pumping still does not prevent to improve the resource availability with respect to the benchmark. This is achieved by the RTOS through a better allocation of pump and

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

release volumes over the optimisation period. When comparing DSP-corr with ESP, we find that the largest improvements with respect to the benchmark are gained in the same years for both products, in the driest years. As already emerged from the analysis of average performances (Figure 4), we see that ESP achieves slightly better resource availability than DSP-corr but with less pumping cost savings. ESP in particular seems to produce 'unnecessary' pumping costs in 2006-07, 2011-12 and 2013-14, where DSP-corr achieves a similar resource availability (top panel) at almost no cost (bottom). It must be noted that for the ESP approach, three specific years play the most important role in decreasing the pumping energy cost savings with respect to the benchmark, 2006-07, 2011-12 and 2013-14 (Figure 7b), which together with 2010-11 have the lowest initial storage (Figure 7a).

4. Discussion

Our study provides some insights on the complex relationship between forecast skill and its value for decision-making. Although these findings may be dependent on the case study and time period that was available for the analysis, they still enable us to draw some more general lessons that could be useful also beyond the specific case investigated here.

First, we found that the use of bias correction to improve the skill and value of DSP forecast is less straightforward than possibly expected. Our results show that on average bias correction slightly improves the DSP forecast skill (as measured by CRPSS and mean error) but it can reduce it in dry years (Figure 3). This is because in our system DSP forecasts systematically underestimate inflows (before bias correction), which means their skill is relatively higher in exceptionally dry years and is deteriorated by bias correction. To our knowledge, no previous study reported such difference in skill for the ECMWF SEAS5 forecasts in dry years in the UK, hence we are not able to say whether our result applies to other systems in the region. However, the result points at a possible intrinsic contradiction in the very idea of bias correcting based on climatology-based forecast (e.g. ESP). In fact, by pushing forecasts to be more alike climatology, one may reduce the 'good signal' that may be present in the original forecast in years that will indeed be significantly drier (or wetter) than climatology. As exceptional conditions are likely the ones when water managers can extract more value from forecasts, the argument that bias correction ensures average performance at least equivalent to ESP (e.g. Crochemore et al. (2016)) may not be very relevant here. We would conclude that more studies are needed to investigate the benefits of bias correction when seasonal hydrological forecasts are specifically used to inform water resource management.

While we could not find an obvious and significant improvement of forecast skill after bias correction, we found a clear increase in forecast value (Figure 4). In fact, RTOS based on bias-corrected DPS considerably reduce pumping costs with respect to original DPS, while ensuring similar resource availability. We explained this finding by the change in the sign of forecasting errors induced by bias correction – from a systematic underestimation of inflows to a systematic overestimation. While this change is again case specific, a general implication is that not all forecast errors have the same impact on the forecast value. From a water resource management perspective, the improvement of forecast accuracy in some directions can be more 'valuable' than others. This also implies that not all skill scores may be equally useful and relevant for water resource managers. For example, in our case a score that is able to differentiate between overestimation and underestimation

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

error, such as the mean error, seems more adequate than a score such as CRPSS, which is insensitive to the error sign. This said, our results overall suggest that inferring the forecast value from its skill may be misleading, given the weak correlation between the two (at least as long as we use skill scores that are not specifically tailored to water resources management). Running simulation experiments of the system operation, as done in this study, can shed more light on the value of different forecast products.

While we found a weak correlation between forecast skill and value, we found that forecast value is more strongly liked to hydrological conditions (Figure 6). As expected, a forecast-based RTOS system is particularly useful in dry years, where we find most of the gains with respect to the benchmark operation (Figure 7). This is consistent with previous studies for water supply system, e.g. Turner et al., 2017. An interesting finding in our system is that the value of forecast-based RTOS seems correlated to the Initial conditions (total storage value) of the system. Given that this initial condition is known at the beginning of the pumped-storage season, in practice this indicator could be used to decide whether to use the forecast-based RTOS approach in the coming months or not. In fact, using the RTOS has a cost in that downloading seasonal weather forecasts, transforming them into hydrological forecasts and bias correcting, running optimisation, etc. takes time. So, water managers may choose to use RTOS only in those years where they expect it will lead to considerable improvements of system performance.

Similarly, in light of the pre-processing costs of seasonal weather forecasts, it is interesting to discuss whether their use is justified with respect to a possibly simpler-to-use product such as ESP. In this study, we found ESP to be a 'hard-to-beat' reference not only in terms of skills (as previously found by others, e.g. (Harrigan et al., 2018)) but also in terms of forecast value (Figure 4). In fact, the use of DSP-corr delivers higher energy savings with respect to ESP (without compromising the resource availability) at least in the most relevant operating priority scenario (the *rap* scenario, see Figure 7). However, whether these cost-savings are large enough to justify the use of DSP-corr, or whether water managers may fall back on using simpler ESP, is difficult to argue and remain an open question with the simulations results available so far.

One point where our results instead point to a univocal and clear conclusion is in the importance of explicitly considering forecast uncertainty (Figure 5). In fact, RTOS outperforms the current operation when using ensemble forecasts, but it does not if uncertainty is removed and the ensemble mean is used within a deterministic optimisation approach. This is in line with previous results obtained using short-term forecasts for flood control (Ficchì et al., 2016), who found that consideration of forecast uncertainty could largely compensate the loss in value caused by forecast errors), hydropower generation (Boucher et al., 2012) and multi-purpose systems (Yao and Georgakakos, 2001). It is also consistent with previous results by Anghileri et al. (2016), who did not find significant value in seasonal forecasts while using a deterministic optimisation approach (they did not explore the use of ensemble though). From the UK water industry perspective, we hope our results will motivate a move away from the deterministic (worst-case scenario) approach that often prevails when using models to support short-term decisions, and a shift towards more explicit consideration of model uncertainties. Such a move would also align with the advocated use of "risk-based" approaches for long-term planning (Hall et al., 2012, Turner et al., 2016, UKWIR, 2016a, UKWIR, 2016b), which have indeed been adopted by water companies in the preparation of their Water

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

Resource Management Plans (SouthernWater, 2018, UnitedUtilities, 2019). The results presented here, and in the above cited studies, suggest that greater consideration of uncertainty and trade-offs would also be beneficial in short-term production planning. Last, we tried to investigate whether we could evaluate the effect of the ensemble size on the value of the uncertain forecasts. We found that in our case study we could reduce the number of forecast members down to about 10 (from the original size of 25) with limited impact on the forecast value (Figure 5). This is important for practice because by reducing the number of forecast members one can reduce the computation time of the RTOS. While we cannot say if such 'optimal' ensemble size would apply to other systems too, we would suggest that future studies could look at how the quality of the uncertainty characterisation impacts on the forecast value, and whether a 'minimum representation of uncertainty' exists that ensures the most effective use of forecasts for water resource management.

4.1 Limitations and perspective for future research and implementation

Our study is subject to a range of limitations that should kept in mind when evaluating our results. First, the current (and future) skill of seasonal forecasts varies spatially across the UK depending on the influence of climate teleconnections and particularly the NAO. Given that our case study is located in the West of the UK, where the NAO influence has been found to be stronger than in the East (Svensson et al., 2015), our simulated benefits of using DSP seasonal forecasts may be particularly optimistic. Second, the general validity of the results is limited by the relatively short period (2005-2016) that was available for historical simulations, and which may be insufficient to fully characterise the variability of hydrological conditions and hence accurately estimate the system's performances (see for example discussion in Dobson et al. (2019)). Hence we aim at continuing the evaluation of the RTOS over time as new seasonal forecasts and observations become available. Another limitation of evaluation of the RTOS is that we used the observed water demand, hence implicitly assuming that operators know in advance the demand values for the entire season with full certainty.

The Python code developed to generate the seasonal inflow forecasts, to optimise the system operation and to visualise the pre-evaluation Pareto front (with its uncertainty), has been implemented in a set of interactive Jupyter Notebooks, which we have now transferred to the water company in charge of the pumped-storage decisions. This toolkit aims at addressing some of the problems identified in the literature for the implementation of forecast informed reservoir operation systems, by providing better "packaging" (Goulter, 1992) of model results and their uncertainties, enabling the interactive involvement of decision makers (Goulter, 1992) and creating a standard and formal methodology (Labadie, 2004) to support model-informed decisions. Besides supporting the specific decision-making problem faced by the water company involved in this study, through this collaboration we aim at evaluating more broadly the effectiveness of our toolkit to promote knowledge transfer from the research to the professional community. Through the use of the toolkit, we also hope to gain a better understanding of how decision-makers view forecast uncertainty, the institutional constraints limiting the use and implementation of this information (Rayner et al., 2005) and the most effective ways in which forecast uncertainty and simulated system robustness can be represented.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

5. Conclusions

This work assessed the potential of using a real-time optimization system informed by seasonal forecasts to improve reservoir operation in a UK water supply system. While the specific results are only valid for the studied system, they enable us to draw some more general conclusions. First, we found that the use of seasonal forecasts can improve the efficiency of reservoir operation, but only if the forecast uncertainty is explicitly considered (e.g. via ensemble forecast). Second, while dynamical streamflow predictions (DSP) generated by numerical weather predictions provided the highest value in our case study (under a scenario that prioritise water availability over pumping costs), still ensemble streamflow predictions (ESP), which are more easily derived from observed meteorological conditions in previous years, remain a hard-to-beat reference in terms of both skill and value. Third, the relationship between the forecast skill and its value for decision-making is complex and strongly affected by the decision maker priorities and the hydrological conditions in each specific year. It must be noted that in practice the decision-making priorities are not solely related to the selection of a specific Pareto-optimal solution, but also the methodology in the first place, i.e. the "risk" taken in using something other than the worst-case scenario and in applying bias correction or not. We hope that this study will contribute to show that seasonal forecasts can deliver benefits to inform operational decisions even if their skill is low; and stimulate further research towards better understanding the skill-value relationship and finding ways to extract value from forecasts in support of water resource management.

Data availability. The reservoir system data used are property of Wessex Water and as such cannot be shared by the authors. ECMWF data are available under a range of licences. For more information please visit http://www.ecmwf.int.

Author contributions. AP developed the model code and performed the simulations under the supervision of FP. CH helped to frame the case study and in the interpretation of the results. All the authors contributed to the writing of the manuscript.

Competing interests. We declare that there are no competing interests.

Acknowledgments. This work is funded by the Engineering and Physical Sciences Research Council (EPSRC), grant EP/R007330/1. The authors are also very grateful to Wessex Water for the data provided. The authors wish to thank the Copernicus Climate Change and Atmosphere Monitoring Services for providing the seasonal forecasts generated by the ECMWF seasonal forecasting systems (SEAS5). Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

References

ALEMU, E. T., PALMER, R. N., POLEBITSKI, A. & MEAKER, B. 2010. Decision support system for optimizing reservoir operations using ensemble streamflow predictions. Journal of Water Resources Planning and Management, 137, 72-82.

ANDERSON, J. L. 1996. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. Journal of climate, 9, 1518-1530.

ANGHILERI, D., VOISIN, N., CASTELLETTI, A., PIANOSI, F., NIJSSEN, B. & LETTENMAIER, D. P. 2016. Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. Water Resources Research, 52, 4209-4225.

ARNAL, L., CLOKE, H. L., STEPHENS, E., WETTERHALL, F., PRUDHOMME, C., NEUMANN, J., KRZEMINSKI, B. & PAPPENBERGER, F. 2018. Skilful seasonal forecasts of streamflow over Europe? Hydrology and Earth System Sciences, 22, 2057.

BAZILE, R., BOUCHER, M. A., PERREAULT, L. & LECONTE, R. 2017. Verification of ECMWF System 4 for seasonal hydrological forecasting in a northern climate. Hydrol. Earth Syst. Sci., 21, 5747-5762.

BELL, V. A., DAVIES, H. N., KAY, A. L., BROOKSHAW, A. & SCAIFE, A. A. 2017. A national-scale seasonal hydrological forecast system: development and evaluation over Britain. Hydrology and Earth System Sciences, 21, 4681.

BERGSTRÖM, S. & SINGH, V. 1995. The HBV model. Computer models of watershed hydrology., 443-476.

BLOCK, P. & RAJAGOPALAN, B. 2007. Interannual Variability and Ensemble Forecast of Upper Blue Nile Basin Kiremt Season Precipitation. Journal of Hydrometeorology, 8, 327-343.

BOUCHER, M. A., TREMBLAY, D., DELORME, L., PERREAULT, L. & ANCTIL, F. 2012. Hydro-economic assessment of hydrological forecasting systems. Journal of Hydrology, 416-417, 133-144.

BROWN, C. M., LUND, J. R., CAI, X., REED, P. M., ZAGONA, E. A., OSTFELD, A., HALL, J., CHARACKLIS, G. W., YU, W. & BREKKE, L. 2015. The future of water resources systems analysis: Toward a scientific framework for sustainable water management. Water Resources Research, 51, 6110-6124.

BROWN, T. A. 1974. Admissible scoring systems for continuous distributions (Report P-5235).

CROCHEMORE, L., RAMOS, M. H. & PAPPENBERGER, F. 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrol. Earth Syst. Sci., 20, 3601-3618.

DAY, G. N. 1985. Extended Streamflow Forecasting Using NWSRFS. Journal of Water Resources Planning and Management, 111, 157-170.

DEB, K., PRATAP, A., AGARWAL, S. & MEYARIVAN, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6, 182-197.

DOBSON, B., WAGENER, T. & PIANOSI, F. 2019. How Important Are Model Structural and Contextual Uncertainties when Estimating the Optimized Performance of Water Resource Systems? 55, 2170-2193.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

EA 2017. Water resources management planning guideline: Interim update.

EHRET, U., ZEHE, E., WULFMEYER, V., WARRACH-SAGI, K. & LIEBERT, J. 2012. HESS Opinions "Should we apply bias correction to global and regional climate model data?". Hydrol. Earth Syst. Sci., 16, 3391-3404.

EPSTEIN, E. S. 1969. A scoring system for probability forecasts of ranked categories. Journal of Applied Meteorology, 8, 985-987.

FABER, B. A. & STEDINGER, J. R. 2001. Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. Journal of Hydrology, 249, 113-133.

FAN, F. M., SCHWANENBERG, D., ALVARADO, R., ASSIS DOS REIS, A., COLLISCHONN, W. & NAUMMAN, S. 2016. Performance of Deterministic and Probabilistic Hydrological Forecasts for the Short-Term Optimization of a Tropical Hydropower Reservoir. Water Resources Management, 30, 3609-3625.

FICCHÌ, A., RASO, L., DORCHIES, D., PIANOSI, F., MALATERRE, P.-O., OVERLOOP, P.-J. V. & JAY-ALLEMAND, M. 2016. Optimal Operation of the Multireservoir System in the Seine River Basin Using Deterministic and Ensemble Forecasts. Journal of Water Resources Planning and Management, 142, 05015005.

GEORGAKAKOS, K. P. & GRAHAM, N. E. 2008. Potential Benefits of Seasonal Inflow Prediction Uncertainty for Reservoir Release Decisions. Journal of Applied Meteorology and Climatology, 47, 1297-1321.

GLEICK, P. H. 2003. Global Freshwater Resources: Soft-Path Solutions for the 21st Century. Science, 302, 1524-1528.

GLEICK, P. H., COOLEY, H., FAMIGLIETTI, J. S., LETTENMAIER, D. P., OKI, T., VÖRÖSMARTY, C. J. & WOOD, E. F. 2013. Improving understanding of the global hydrologic cycle. Climate science for serving society. Springer.

GOULTER, I. C. 1992. Systems Analysis in Water‐Distribution Network Design: From Theory to Practice. Journal of Water Resources Planning and Management, 118, 238-248.

GREUELL, W., FRANSSEN, W. H. P. & HUTJES, R. W. A. 2019. Seasonal streamflow forecasts for Europe – Part 2: Sources of skill. Hydrol. Earth Syst. Sci., 23, 371-391.

HADKA, D. 2018. A Free and Open Source Python Library for Multiobjective Optimization [Online]. Available: https://github.com/Project-Platypus/Platypus [Accessed 06/11/2019].

HAGEMANN, S., CHEN, C., HAERTER, J. O., HEINKE, J., GERTEN, D. & PIANI, C. 2011. Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. 12, 556-578.

HALL, J. W., WATTS, G., KEIL, M., DE VIAL, L., STREET, R., CONLAN, K., O'CONNELL, P. E., BEVEN, K. J. & KILSBY, C. G. 2012. Towards risk-based water resources planning in England and Wales under a changing climate. 26, 118-129.

HARRIGAN, S., PRUDHOMME, C., PARRY, S., SMITH, K. & TANGUY, M. 2018. Benchmarking ensemble streamflow prediction skill in the UK. Hydrology and Earth System Sciences, 22, 2023.

HEMRI, S., SCHEUERER, M., PAPPENBERGER, F., BOGNER, K. & HAIDEN, T. 2014. Trends in the predictive performance of raw ensemble weather forecasts. 41, 9197-9205.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

HERSBACH, H. 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. Weather and Forecasting, 15, 559-570.

LABADIE, J. W. 2004. Optimal Operation of Multireservoir Systems: State-of-the-Art Review. Journal of Water Resources Planning and Management, 130, 93-111.

MACLACHLAN, C., ARRIBAS, A., PETERSON, K., MAIDENS, A., FEREDAY, D., SCAIFE, A., GORDON, M., VELLINGA, M., WILLIAMS, A. & COMER, R. 2015. Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. Quarterly Journal of the Royal Meteorological Society, 141, 1072-1084.

MASON, I. 1982. A model for assessment of weather forecasts. Aust. Meteor. Mag, 30, 291-303.

MAURER, E. P. & LETTENMAIER, D. P. 2004. Potential Effects of Long-Lead Hydrologic Predictability on Missouri River Main-Stem Reservoirs. Journal of Climate, 17, 174-186.

PAPPENBERGER, F., RAMOS, M. H., CLOKE, H. L., WETTERHALL, F., ALFIERI, L., BOGNER, K., MUELLER, A. & SALAMON, P. 2015. How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. Journal of Hydrology, 522, 697-713.

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. & JENKINS, A. 2017. Hydrological Outlook UK: an operational streamflow and groundwater level forecasting system at monthly to seasonal time scales. Hydrological Sciences Journal, 62, 2753-2768.

RAYNER, S., LACH, D. & INGRAM, H. 2005. Weather forecasts are for wimps: why water resource managers do not use climate forecasts. Climatic Change, 69, 197-227.

SCAIFE, A. A., ARRIBAS, A., BLOCKLEY, E., BROOKSHAW, A., CLARK, R. T., DUNSTONE, N., EADE, R., FEREDAY, D., FOLLAND, C. K., GORDON, M., HERMANSON, L., KNIGHT, J. R., LEA, D. J., MACLACHLAN, C., MAIDENS, A., MARTIN, M., PETERSON, A. K., SMITH, D., VELLINGA, M., WALLACE, E., WATERS, J. & WILLIAMS, A. 2014. Skillful long-range prediction of European and North American winters. Geophysical Research Letters, 41, 2514-2519.

SOUTHERNWATER. 2018. Revised draft Water Resources Management Plan 2019 Statement of Response [Online]. Available: https://www.southernwater.co.uk/media/1884/statement-of-response-report.pdf [Accessed 6/11/19 2019].

SVENSSON, C., BROOKSHAW, A., SCAIFE, A. A., BELL, V. A., MACKAY, J. D., JACKSON, C. R., HANNAFORD, J., DAVIES, H. N., ARRIBAS, A. & STANLEY, S. 2015. Long-range forecasts of UK winter hydrology. Environmental Research Letters, 10, 064006.

TIM, S., MAGDALENA, A.-B., STEPHANIE, J., LAURA, F., FRANCO, M., MAGNUSSON, L., STEFFEN, T., FRÉDÉRIC, V., DAMIEN, D., ANTJE, W., CHRISTOPHER, D. R., GIANPAOLO, B., SARAH, K., KRISTIAN, M., HAO, Z., MICHAEL, M. & MONGE-SANZ, B. M. 2018. SEAS5 and the future evolution of the long-range forecast system. ECMWF.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020


Hydrology and Earth System Sciences
Discussions

TURNER, S. W. D., BENNETT, J. C., ROBERTSON, D. E. & GALELLI, S. 2017. Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. Hydrol. Earth Syst. Sci., 21, 4841-4859.

TURNER, S. W. D., BLACKWELL, R. J., SMITH, M. A. & JEFFREY, P. J. 2016. Risk-based water resources planning in England and Wales: challenges in execution and implementation. Urban Water Journal, 13, 182-197.

UKWIR 2016a. WRMP 2019 Methods - Decision making process: Guidance.

UKWIR 2016b. WRP19 Methods – Risk-Based Planning.

UNITEDUTILITIES. 2019. Final water resources management plan 2019 [Online]. Available: https://www.unitedutilities.com/globalassets/z_corporate-site/about-us-pdfs/wrmp-2019---2045/final-water-resources-management-plan-2019.pdf [Accessed 6/11/19 2019].

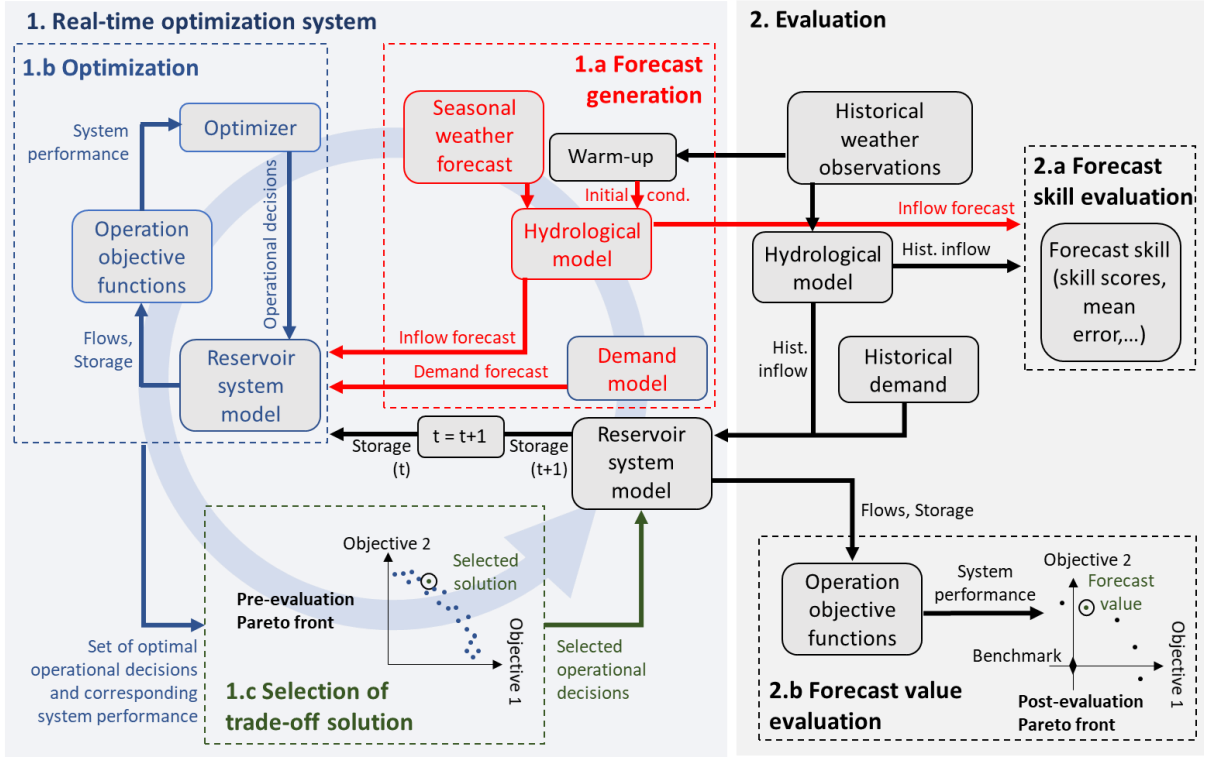
VOISIN, N., PAPPENBERGER, F., LETTENMAIER, D. P., BUIZZA, R. & SCHAAKE, J. C. 2011. Application of a Medium-Range Global Hydrologic Probabilistic Forecast Scheme to the Ohio River Basin. Weather and Forecasting, 26, 425-446.

WANG, F., WANG, L., ZHOU, H., SAAVEDRA VALERIANO, O. C., KOIKE, T. & LI, W. 2012. Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. Water Resources Research, 48.

WANG, L., TING, M. & KUSHNER, P. J. 2017. A robust empirical seasonal prediction of winter NAO and surface climate. Scientific Reports, 7, 279.

YAO, H. & GEORGAKAKOS, A. 2001. Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management. Journal of Hydrology, 249, 176-196.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020



620 **Figure 1 Diagram of the methodology used in this study to generate operational decisions using a Real-time optimisation system (RTOS) (left) and to evaluate its performances (right). In the evaluation step, the RTOS is nested into a closed loop simulation where at every time step historical data (weather, inflows and demand), along with the operational decisions suggested by the RTOS, are used to move to the next step by updating the initial hydrological conditions and reservoir storage.**

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020

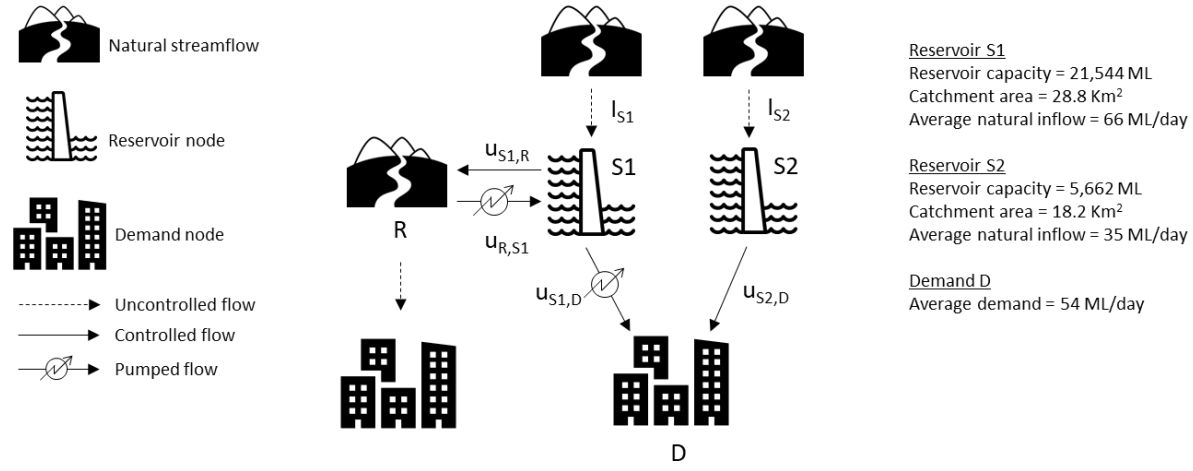
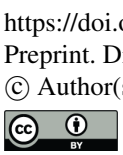


Figure 2 A schematic of the reservoir system investigated in this study to test the Real-time optimization systems. I is natural reservoir inflow, S reservoir node, u controlled inflows/releases, R river and D human demand node. The system is a two-reservoir system where S1 both supports downstream abstraction during low river (R) flows and use pumped releases to complement gravity releases from S2 in supplying D. The system has the possibility of pumping water into S1 from Nov to Apr.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020



Hydrology and Earth System Sciences
Discussions

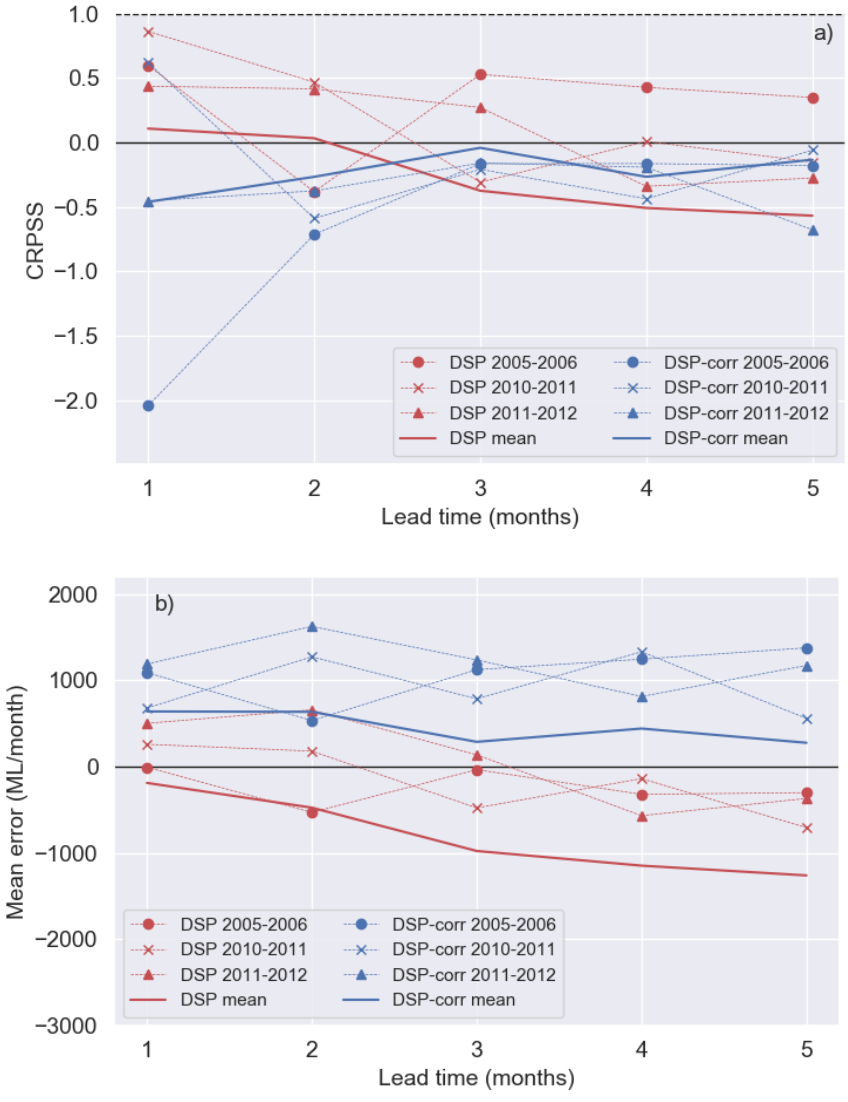
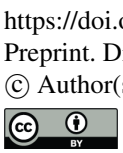


Figure 3 Skill of the hydrological forecast ensemble measured by the CRPSS (a) and the mean error (b) for different lead times. Red lines represent the skill of the non-bias corrected ECMWF seasonal forecast, blue lines represent the skill after bias correction. The solid line represents the mean skill over the period 2005-2016, while circles, crosses and triangles represent the skill in 3 particularly dry winters. CRPSS = 1 represents the perfect forecast and CRPSS = 0 the no skill threshold with respect to the benchmark (ESP).

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

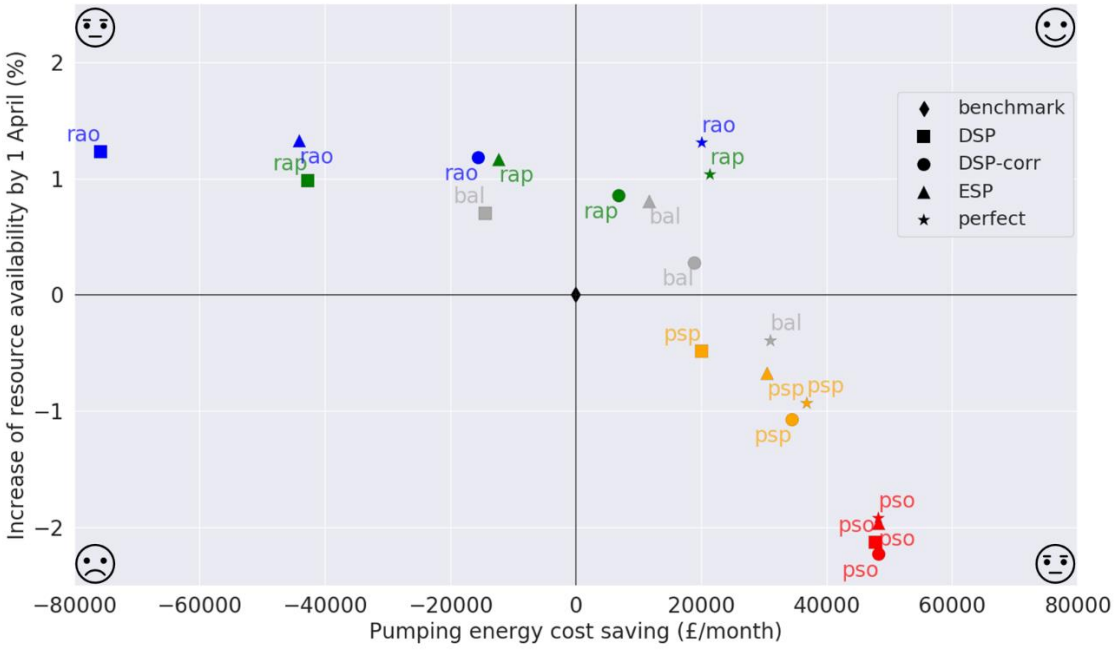


Figure 4 Post-evaluation Pareto fronts representing the average system performance improvement (over period 2005-2016) of the real-time optimization system during the pumping licence window (1 Nov - 1 Apr) with respect to the benchmark (black diamond), using four forecast products: non-corrected forecast ensemble (DSP), bias corrected forecast ensemble (DSP-corr), ensemble streamflow prediction (ESP) and perfect forecast. For each of the four forecast products five decision making priorities are represented depending on the dominant priority from 100% priority to maximize resource availability (top left) to 100% priority to maximize cost savings (bottom right): resource availability only (rao; in blue), resource availability prioritised (rap; in green), balanced (bal; in grey), pumping savings prioritised (psp; in green) and pumping savings only (pso; in red). The pumping energy cost is calculated as the sum of the energy costs associated to pumped inflows and pumped releases and the resource availability as the mean storage volume in both reservoirs (S1 and S2) at the end of the optimisation period. The annotation is the corresponding operational priority scenario for each point.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

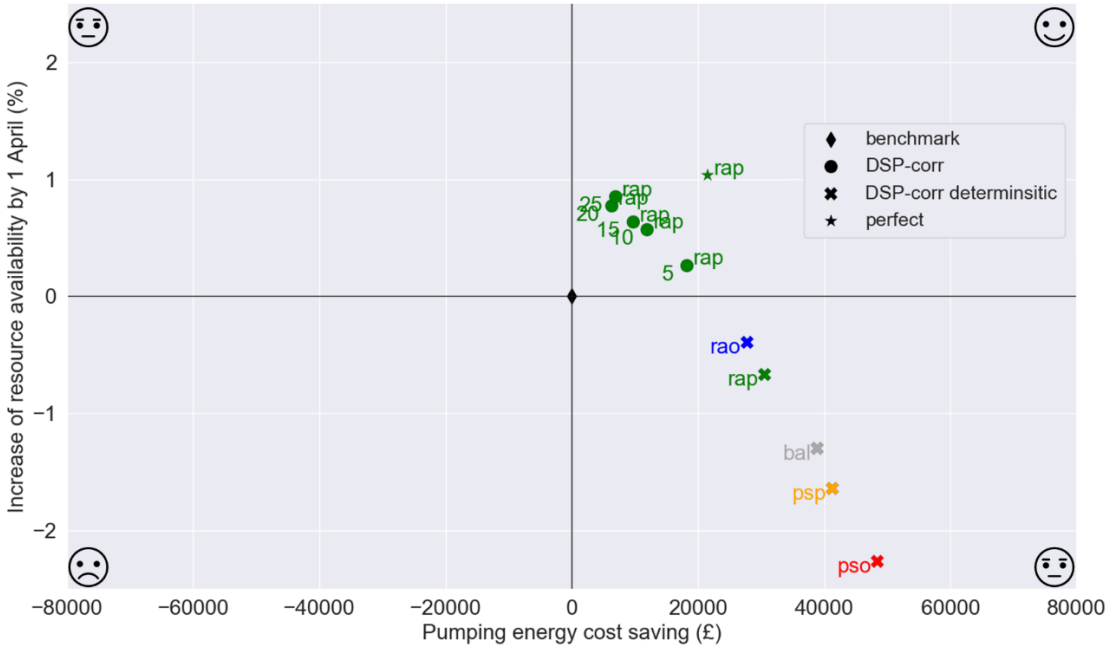


Figure 5 Post-evaluation Pareto fronts representing the average system performance (over period 2005-2016) of the real-time optimization system during the pumping licence window (1 Nov - 1 Apr) with respect to the benchmark (black diamond), using bias corrected forecast deterministic (DSP-corr deterministic) and bias corrected forecast ensemble (DSP-corr) with different ensemble size. For practical purposes, only the "resource availability prioritised" scenario (rap) is represented for the DSP-corr. The annotation is the corresponding operational priority for each point. The annotation numbers are the number of ensemble members considered. The perfect forecast for rap is also represented for reference purposes.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020

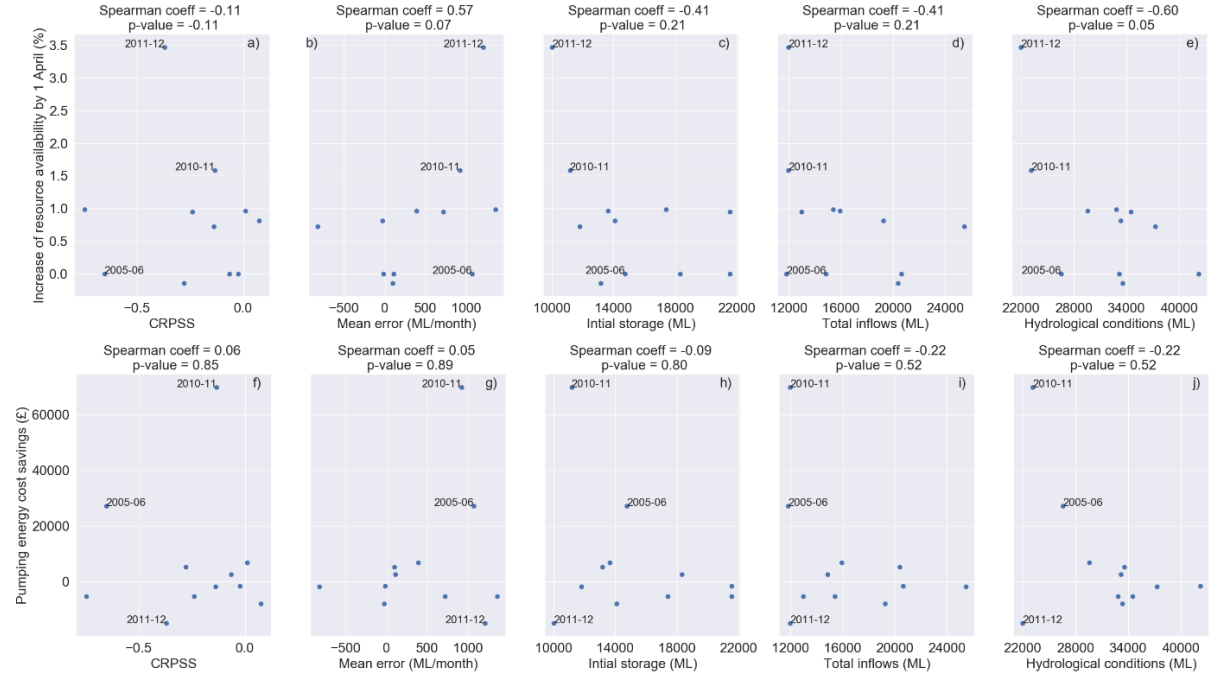


Figure 6 Bias corrected forecast ensemble (DSP-corr) for the "resource availability prioritised" scenario - correlation between Increase of resource availability and a) CRPSS, b) mean error, c) initial storage (1 Nov), d) total inflows (1 Nov – 1 Apr) and e) hydrological conditions (initial storage + total inflows) and between Pumping energy cost savings and f) CRPSS, g) mean error, h) initial storage (1 Nov), i) total inflows (1 Nov – 1 Apr) and j) hydrological conditions (initial storage + total inflows). Each point represents a year. Correlation and its significance are quantified by the Spearman coefficient and the p-value, respectively.

https://doi.org/10.5194/hess-2020-89
Preprint. Discussion started: 6 March 2020

Hydrology and Earth System Sciences
Discussions

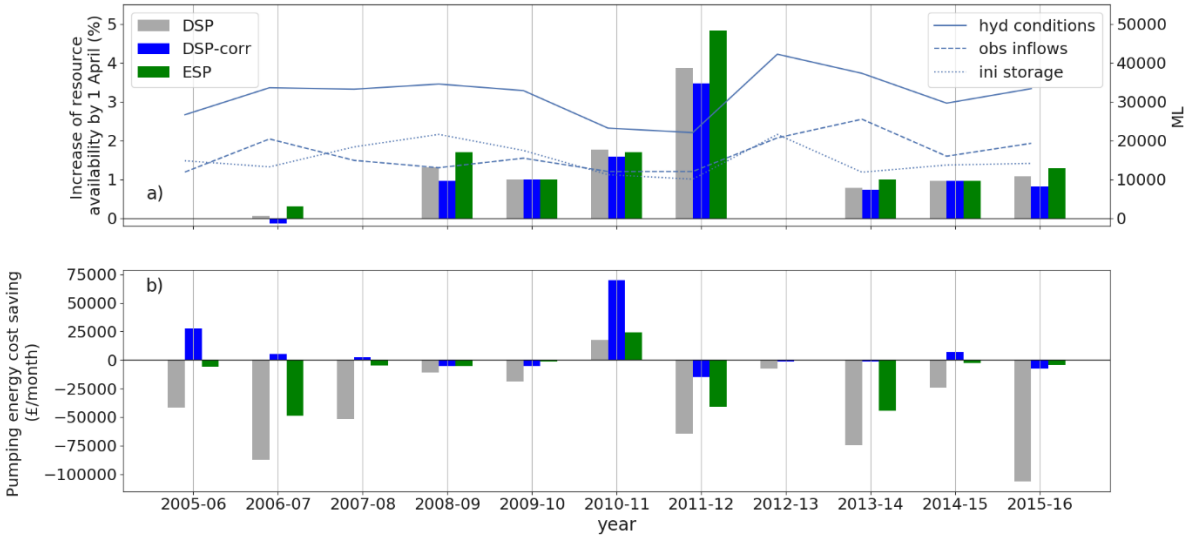


Figure 7 Year-by-year a) Total observed inflows (1 Nov - 1 Apr), Initial reservoir storage (1 Nov) and Hydrological conditions (Total observed inflows + Initial storage) (right hand y-axis) and Increase of resource availability (left hand y-axis) and b) Pumping energy cost savings of the real operation system informed by: the dynamical streamflow prediction (DSP), the bias corrected dynamical streamflow prediction (DSP-corr) and the ensemble streamflow prediction (ESP) for the "resource availability prioritised" (rap) scenario.