Unfolding the relationship between seasonal forecasts skill and value in hydropower production: A global analysis

Donghoon Lee[1], Jia Yi Ng[2], Stefano Galelli[2], and Paul Block[1]

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2]Pillar of Engineering Systems and Design, Singapore University of Technology and Design, Singapore

Key Points:

- 25% of the 735 headwater hydropower dams evaluated worldwide may benefit from seasonal forecasts conditioned on hydroclimatic predictors.
- Potential benefits are predominantly modulated by forecast skill and reservoir characteristics.
- We identify geographical regions where dams would benefit the most from forecasts

Corresponding author: Stefano Galelli, stefano_galelli@sutd.edu.sg

Abstract

The potential benefits of seasonal streamflow forecasts for the hydropower sector have been evaluated for several basins across the world, but with contrasting conclusions on expected hydropower production and economic gains. This raises the prospect of a complex relationship between reservoir characteristics, forecast skill and value. Here, we unfold the nature of this relationship by studying time series of simulated power production for 735 headwater dams worldwide. The time series are generated by running a detailed dam model over the period 1958-2000 with three operating schemes: basic control rules, perfect forecast-informed, and realistic forecast-informed. The realistic forecasts are issued by bespoke models, based on lagged global and local hydroclimatic variables, predicting seasonal monthly dam inflows. Results show that most dams (94%) could benefit from perfect forecasts. Yet, the benefits for each dam vary greatly and are primarily controlled by the time to fill and the ratio between reservoir depth and hydraulic head. When realistic forecasts are adopted, 25% of dams demonstrate improvements with respect to basic control rules. In this case, the likelihood of observing improvements is controlled not only by design characteristics but also by forecast skill. We conclude our analysis by identifying two groups of dams of particular interest: dams that fall in regions expressing strong forecast accuracy and have the potential to reap benefits from forecast-informed operations, and dams with strong potential to benefit from forecast-informed operations but lack forecast accuracy. Overall, these results represent a first qualitative step towards informing site-specific hydropower studies.

Plain Language Summary

Seasonal streamflow forecasts are an important asset for hydropower operators. Their value has been assessed in several regions, but with contrasting conclusions on how predictive accuracy, or skill, and dam design specifications affect the expected increase in power production. Here, we discover the nature of this relationship by studying a large dataset comprising seasonal forecasts and simulated hydropower production for 735 headwater dams worldwide, representing 10% of the world's installed hydropower capacity. Our results show that 25% of these dams demonstrate improvements. We conclude the analysis by identifying the values of forecast skill and design specifications that are necessary to reap immediate benefits from forecast-informed operations. Overall, the information revealed by this study could support the design and operations of large-scale hydropower projects.

1 Introduction

Hydropower is the leading form of renewable power, contributing to 16% of global electricity production and supplying 62% of all renewable electricity (IHA, 2019). Total hydropower generation is expected to double by 2050, with substantial growth in Asia, Africa, and South America (Zarfl et al., 2015; X. Zhang et al., 2018). Sustainable operations of hydropower facilities, however, are challenged by climate variability and change, which modify short-term and long-term water availability, often with direct effects on regional and global economies (Turner, Hejazi, et al., 2017).

Many studies assessed the potential impacts of climate change on global, continental, and regional hydropower production using projected streamflow from hydrological models (Hamududu & Killingtveit, 2012; Van Vliet et al., 2016; Turner, Ng, & Galelli, 2017; X. Zhang et al., 2018). For example, T. Zhou et al. (2018) outlines expected substantial seasonal changes in hydropower generation in the western United States, while Kao et al. (2015) estimates that the United States federal hydropower production will decrease 1-2 TWh per year until 2039. Decreases in hydropower production in the late 21^{th} century are also expected in Europe (Lehner et al., 2005) and China (Liu et al., 2016), although regional variations are likely. Such projections in hydropower production can prompt policymakers to engage in strategic adaptation, including infrastructure expansion or alternative reservoir operating policies, to ensure water-energy security in the long run (Payne et al., 2004; Van Vliet et al., 2016).

In contrast to climate change, climate variability presents a fundamentally different challenge, namely seasonal and inter-annual fluctuations in streamflow and hydropower output driven by large-scale climate drivers. Examples include the North Atlantic Oscillation (NAO), affecting hydropower in Europe (De Felice et al., 2018), and the El Niño Southern Oscillation (ENSO), affecting one third of the world's hydropower dams (Ng et al., 2017). In theory, the negative impact of climate variability on hydropower production can be tackled with adaptive reservoir operating policies based on seasonal streamflow forecasts, but, in practice, the ripped benefits depend on the complex relationship between hydropower production, forecast skill, and reservoir characteristics.

In previous studies, this relationship has been studied with either analytical or experimental approaches. In the analytical approach, one typically uses synthetic forecasts and hypothetical reservoir systems (e.g., concave objective function, monotonic relationship between current operation decision and ending storage) to analytically derive a relationship between the aforementioned variables. For example, You and Cai (2008) derive a theoretical relationship linking the ideal forecast horizon to various factors, such

as water stress level, reservoir size, or inflow uncertainty. In a follow-up study, Zhao et al. (2012) investigate the relationship between forecast horizon and uncertainty, identifying an effective forecast horizon that balances the effects of horizon and uncertainty, providing the largest benefit to the reservoir operators. In contrast, the experimental approach simulates the operations of existing reservoirs systems with seasonal streamflow forecasts to determine their potential value—and, where possible, to build an empirical relationship linking forecast value, skill, and reservoir characteristics. A common, and expected, conclusion shared by many studies is that incorporating streamflow forecasts into reservoir operating policies can lead to increased hydropower production and economic gains (Kim & Palmer, 1997; Ahmad & Hossain, 2019). What is perhaps more interesting is that the expected gains vary widely. Maurer and Lettenmaier (2004), for instance, observed a modest 1.8% hydropower benefit for reservoirs along the Missouri River utilizing perfect forecasts. They attribute the relatively low gains to the system's large storage capacity relative to annual inflow. Similarly, Rheinheimer et al. (2016) noted an expected 1.2% economic gain for hydropower systems in the Sierra Nevada (California) and found that forecast value is insensitive to storage capacity, yet highly sensitive to powerhouse capacity. By contrast, Hamlet et al. (2002) used ENSO and the Pacific Decadal Oscillation (PDO) signals to construct long-range streamflow forecasts, and estimated that such forecasts could increase hydropower revenue by \$153 million/year ($> 40\%$) for the Columbia River system. Recently, Anghileri et al. (2019) applied subseasonal hydrometeorological forecasts to improve both revenue and unproductive spill for the Verzasca hydropower system in the Swiss Alps. Similar benefits have been demonstrated in many other countries, such as Ecuador (Gelati et al., 2014), Ethiopia (Block, 2011), and the Philippines (Sankarasubramanian et al., 2009; Libisch-Lehner et al., 2019).

While these studies illustrate the potential benefits of seasonal forecasts, they are limited to individual dams or specific river basins. A 'synoptic' assessment of the value of seasonal forecasts for global hydropower production is lacking. In addition, there is only fragmented knowledge on how forecast skill and reservoir characteristics translate into forecast value. Characterizing such relationship across multiple geographic areas and climatic conditions may provide valuable insights for planning and managing hydropower projects. Here, we address these gaps by presenting a global analysis carried out on 753 headwater dams, representing 10% of the world's installed hydropower capacity. Specifically, we leverage recent studies demonstrating global streamflow predictability conditioned on large-scale climate variability (Ward et al., 2014; Lee et al., 2018) and develop seasonal inflow forecasts for each dam. Then, we quantify the value of these forecasts by comparing the amount of hydropower simulated by three operating schemes based on realistic forecasts (issued by our model), perfect forecasts, and (no forecasts) control

rules. With this information at hand, we bank on the wide range of climatic conditions and dam characteristics available in our database to (1) explain how reservoir design properties and forecast skill affect the value of seasonal forecasts, and (2) identify key geographical regions where dams would benefit the most from forecasts. As we shall see, the relationship (between forecast skill and value) and spatial patterns revealed by our analyses represent a first qualitative step towards informing site-specific studies.

2 Data

2.1 Hydropower dams data

We use the database introduced by Ng et al. (2017), which contains design specifications for 1,593 hydropower reservoirs—representing almost 40% of the world's installed hydropower capacity. The database provides information on dam height, storage capacity, maximum surface area, long-term average discharge, upstream catchment area, geographic coordinates, installed power capacity, maximum turbine flow, and operating goals (e.g., hydropower supply, flood control). The majority of these data were originally retrieved from the Global and Dam (GRanD) database (Lehner et al., 2011), and complemented with data from the International Commission on Large Dams (ICOLD, 2011), the Global Lakes and Wetlands Database (Lehner & Döll, 2004), and the Global Energy Observatory (GEO, 2016). The fact that not all dams in the database are located in the headwaters is a major challenge, since the inflows are forecasted based on hydrometeorological data only (see Section 3.1. For this reason, we filter out all dams affected by upstream regulation, reducing the number of dams from 1,593 to 753. To this purpose, we first retrieve data on the Degree Of Regulation (DOR) for each dam, defined as the ratio between the storage volume of the upstream dam(s) and the natural average discharge volume of a given river segment (Grill et al., 2019). We then keep only the dams with DOR equal to 0.

To model the relationship between storage and depth, one ideally needs data on the bathymetry of each reservoir, an information not available at the global scale. Recent advances in remote sensing have shown that these data can be estimated from satellite images (Gao et al., 2012; Bonnema & Hossain, 2017), but for a number of reservoirs that is not yet compatible with the scale of our work (Busker et al., 2019). For this reason, we adopt a simpler bathymetry commonly adopted in global studies (Van Beek et al., 2011; Turner, Ng, & Galelli, 2017). Specifically, we model the storage-depth relationship with Kaveh's method, which assumes an archetypal reservoir shape (Kaveh et al., 2013). This method estimates the reservoir surface area as a function of volume, maximum surface area, depth, and maximum depth. For the limited number of cases in which

155 data on the maximum depth are not available, we adopt Liebe's method, which assumes
156 that a reservoir is shaped like an inverted pyramid cut diagonally in half (Liebe et al.,
157 2005).

158 For each dam, we obtain a monthly inflow time series from the Water and Global
159 Change (WATCH) 20^{th} century model gridded global runoff dataset (Weedon et al., 2011).
160 The runoff data are generated by the global hydrological model WaterGAP (Alcamo et
161 al., 2003), which estimates the accumulated runoff for each grid ($0.5° \times 0.5°$ resolution)
162 using the DDM30 river network (Döll & Lehner, 2002). The model has found success-
163 ful application in various global water resources studies (Döll et al., 2009; Haddeland et
164 al., 2014), but its spatial resolution may be a source of uncertainty for dams located in
165 small catchments. For this reason, we modify the original WATCH database in three ways.
166 First, we consider only the period 1958–2000, which contains more detailed forcing data
167 (Weedon et al., 2011). Second, we manually adjust the position of 270 dams (among the
168 753 dams with DOR equal to 0) to properly align them with the DDM30 river network.
169 To this purpose, we use the HydroSHEDS river network (Lehner et al., 2008) and satel-
170 lite images. Lastly, we correct the discharge data to account for any disparity between
171 the upstream catchment area defined by the DDM30 river network and the documented
172 upstream catchment area of each dam (Ng et al., 2017).

173 In addition to the inflow time series, we also retrieve information on the climate
174 classification of each dam location, an information needed in the latter part of our anal-
175 ysis to characterize the regions that would benefit the most from forecasts. To this pur-
176 pose, we use the updated Köppen-Geiger Climate classification developed on the basis
177 of a large global data set of long-term monthly precipitation and temperature time se-
178 ries (Peel et al., 2007). Specifically, we use the most frequent Köppen-Geiger Climate
179 classification in all upstream grids for each dam.

180 **2.2 Hydro-climatological data**

181 The seasonal forecasts developed here depend on seven predictors: four large-scale
182 climate drivers (ENSO, PDO, NAO, and Atlantic Multidecadal Oscillation (AMO)), and
183 three variables accounting for local processes (lagged inflow, snowfall, and soil moisture.)
184 The four large-scale climate drivers are interannual, decadal, or multidecadal quasiperi-
185 odic oscillations derived from oceanic and atmospheric fields, and play a key role in de-
186 termining climate patterns across the world. To characterize ENSO, we use the Niño 3.4
187 index, defined as the anomalies of 3-month running mean of Sea Surface Temperature
188 (SST) in the Niño 3.4 region (`https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/`
189 `Nino34/`). The monthly PDO index is defined as the leading principal component of monthly

SST anomalies in the North Pacific basin (Y. Zhang et al., 1997). It is obtained from the Joint Institute for the Study of the Atmosphere and Ocean (`http://research.jisao.washington.edu/pdo/`). For NAO, we use the station-based seasonal NAO index, which is the difference in normalized sea level pressure between Reykjavik and Lisbon stations (Hurrell & Deser, 2010) (`https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based`). Finally, the AMO index is defined as the area-weighted average SST over the North Atlantic basin (Enfield et al., 2001). We use the monthly de-trended and un-smoothed AMO index derived from the Kaplan SST (`https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/AMO`). For the PDO and AMO indices, we calculate the 3-month running mean to maintain seasonal persistence.

Monthly soil moisture and snowfall data are obtained from the ERA-40 reanalysis, developed by the European Centre for Medium-Range Weather Forecasts (`https://apps.ecmwf.int/datasets/`) and WATCH forcing data, respectively. For soil moisture, we aggregate all four volumetric soil water layers of the ERA-40. To properly account for the basin-scale soil moisture and snowfall states (Maurer & Lettenmaier, 2004), we calculate the area-weighted average soil moisture and snowfall of all upstream grids for each dam using the DDM30 river network.

3 Methods

The purpose of this study is to (1) quantify the value of seasonal inflow forecasts for global hydropower production, (2) explain how reservoir design properties and forecast skill affect the value of seasonal forecasts, and (3) identify regions that would benefit the most from seasonal forecasts. To achieve these goals, we first develop an inflow prediction model for each of the 753 dams (Section 3.1). Then, we simulate hydropower production for each dam under three operating schemes that are based on perfect forecasts, realistic forecasts (issued by our inflow prediction model), and (no forecast) control rules (Section 3.2). Finally, we evaluate the performance of each operating scheme and identify the reservoir design specifications that explain system's performance (Section 3.3).

3.1 Dam inflow prediction model

Our long-range inflow prediction models are based upon the methodology presented by Lee et al. (2018), who employed lagged large-scale climate drivers and prior streamflow conditions to predict streamflow at 1,200 stations globally. Lee et al. (2018) suggested that a Principal Component Regression (PCR) model with a set of predictors can

provide fair (realistic) predictive skills that can also be easily implemented globally. While Lee et al. (2018) predicted the seasonal (3-month) streamflow, here we develop independent monthly prediction (MP) models for the subsequent seven calendar months. For example, at the end of February, we predict monthly inflows from March (MP1) to September (MP7).

The methodology relies on the following steps, illustrated in Figure 1. First, we normalize (log-normalize for streamflow) and detrend all predictors and streamflow observations to avoid artificial skill due to potential dependence. Then, we estimate the lag-correlations between monthly inflows over the next 7 months and climate indices (1-8 months ahead), snowfall (current to 8 months ahead), and inflow and soil moisture (current month). Statistically significant predictors are subsequently used to develop the MP models. If a single (statistically significant) predictor exists, we apply a linear regression (LR) model; otherwise, we apply the PCR model to avoid possible multicollinearities. In the PCR process, we truncate only the last principal component, which is associated with multicollinearities, as suggested by Jolliffe (2002) and Wilks (2011). We apply a leave-one-out cross-validation (LOOCV) scheme to select the optimal lead-times of the lagged predictors. Specifically, all combinations of lead months for the lagged predictors are cross-validated with a block size of 3 years; then, the optimal set of lead-months is determined based on the minimum mean squared error (MSE). The models are calibrated with 70% of the available data (corresponding to the period 1958–1987) and validated with the remaining data (1988–2000). In the validation process, we evaluate the model performance using two skill scores, namely the mean squared error skill score (MSESS) and the Gerrity skill score (GSS) (Appendix A), as in Lee et al. (2018). If an MP model has no statistically significant predictors, or either an MSESS or GSS value less than 0, the climatological mean prediction is used instead.

The overall accuracy of the reservoir inflow predictions is assessed with the Kling-Gupta efficiency (KGE), which compares correlation, bias, and variability of the predicted and observed discharge (Gupta et al., 2009). The KGE is defined as:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \tag{1}$$

where r is the correlation coefficient, β the bias ratio of the mean inflow (μ_s/μ_o), γ the variability ratio (CV_s/CV_o), μ the mean flow, CV the coefficient of variation, and s and o two indices indicating simulated (predicted) and observed inflow values, respectively.

3.2 Reservoir operation model

3.2.1 Reservoir model

An essential component of the operating schemes described below is the reservoir mass balance:

$$S_{t+1} = S_t + Q_t - E_t - R_t - Spill_t, \tag{2a}$$

$$0 \leq S_t \leq S_{cap}, \tag{2b}$$

$$0 \leq R_t \leq \min(S_t + Q_t - E_t, R_{max}), \tag{2c}$$

where S_t is the reservoir storage at month t, Q_t the inflow volume (retrieved from the WaterGAP model, as described in Section 2.1), E_t the evaporation loss, and R_t the water released through the turbines. Both S_t and R_t are constrained by the reservoir design specifications. Specifically, the storage cannot exceed the reservoir capacity S_{cap} (eq. (2b)), while the discharge is bounded by the water availability and capacity R_{max} of the turbines (eq. (2c)). Excess water, if any, is spilled:

$$Spill_t = \max(0, S_t + Q_t - R_t - E_t - S_{cap}). \tag{2d}$$

The hydropower production P_t (in MW) is calculated as follows:

$$P_t = \eta \cdot \rho \cdot g \cdot r_t \cdot h_t, \tag{3}$$

where η is the efficiency of the turbines assumed constant over the simulation period), ρ the water density (1,000 kg/m^3), g the gravitational acceleration (m/s^2), r_t the average release rate (m^3/s) implied by the monthly release volume R_t, and h_t the hydraulic head (m). The latter is taken as the average head between time t and $t+1$.

3.2.2 Benchmark scheme: control rules

Our benchmark operating scheme relies on the approach proposed by Ng et al. (2017), in which the behaviour of the hydropower operators is modelled as an optimal control problem. This approach builds on two main assumptions, on which we shall return in Section 5. First, the goal of the operators is to maximize hydropower production over

the long term. This objective provides a tangible indication of hydropower performance, so it is commonly adopted in large-scale studies (e.g., Van Vliet et al. (2016)). Second, the release decision R_t depends on the reservoir storage S_t, the previous period's inflow volume Q_{t-1}, and month of year t—a common choice in real-world reservoir operating schemes (Hejazi et al., 2008). In other words, the approach assumes that each reservoir is operated through a bespoke, periodic look-up table of turbine release decisions, which is generated with stochastic dynamic programming (Loucks et al., 2005; Soncini-Sessa et al., 2007). In the optimization, the inflow process is modelled with a first order, periodic Markov chain, whose parameterization is derived from the inflow data. A detailed validation of the operating rules—based on values of observed hydropower production in 107 countries during the period 1980–2000—is reported in Turner, Ng, and Galelli (2017). The time series of all process variables (e.g., inflow, storage, release, hydropower production) obtained by the benchmark control scheme are available on HydroShare (`http://www.hydroshare.org/resource/ca365ffb1a1f49df8b77e393be965fd8`).

3.2.3 Forecast-informed scheme

To assess the value of seasonal streamflow forecasts, we adopt an adaptive scheme based on the *receding horizon principle* (Bertsekas, 1976): at month t, we use a 7-month streamflow forecast to determine the value of the release decisions for the next seven months, and then implement only the decision R_t for the first month. At month $t+1$, when a new 7-month forecast becomes available, a new sequence of release decisions is determined. Each decision-making process is formulated through an optimization problem that maximizes the hydropower production over the forecast horizon while accounting for the benefits associated with the resulting storage at the end of the forecast horizon:

$$\min_{R_t, R_{t+1}, \ldots, R_{t+6}} \sum_{i=0}^{6} P_{t+i} + X(S_{t+7}) \tag{4}$$

where P_t is the hydropower production (see eq. (3)) and $X(\cdot)$ a function accounting for the long-term effect of the release decisions. Specifically, the function penalizes decisions that solely optimize energy production in the short term, risking depleted water availability in the long term. Following a common practice in forecast-informed schemes (Soncini-Sessa et al., 2007), we set $X(\cdot)$ equal to the benefit function obtained by the benchmark control rules, which contains information about the expected long-term hydropower production for a given storage level. Thus, the real-time information provided by the forecasts may alter decisions otherwise based solely on the benchmark scheme (Turner, Bennett, et al., 2017). The optimization problem is solved at each time step using deterministic dynamic programming [ibidem].

The scheme is implemented using both 'perfect' and realistic forecasts (described in Section 3.1). Both benchmark and forecast-informed schemes are simulated over the period 1958–2000. During the simulation, all release decisions are constrained to satisfy downstream environmental flow requirements, calculated using the variable monthly flow method (Pastor et al., 2014). All experiments are carried out with the R package *reservoir* (Turner & Galelli, 2016).

3.2.4 Dealing with additional operating objectives and finer temporal scales

Of the 735 dams, 174 dams within the database are also operated for flood control purposes. For these dams, we penalize spill to account for flood control and formulate the optimization objective as follows (in both benchmark and forecast-informed schemes):

$$\min \sum_{t} (w_1 \cdot \frac{Spill_t}{p_{95}(Q)} + w_2 \cdot (1 - \frac{P_t}{P})) \tag{5}$$

where w_1 and w_2 are the weights associated to the flood control and hydropower objectives, set to 0.5 here, $p_{95}(Q)$ the 95^{th} percentile of the inflow time series Q, and P the installed hydropower capacity (in MW). The presence of an additional goal may result in a change of the hydraulic head or release trajectory, thereby affecting hydropower production (Zeng et al., 2017).

A second modification of the reservoir operation model concerns the monthly decision-making time step, which may not be suitable for reservoirs with small storage capacity relative to inflow (or time-to-fill). We therefore identify a group of 94 reservoirs for which the time-to-fill is smaller than two months, and adopt for this group only a weekly time step. Since the inflow forecasts have a monthly resolution, we disaggregate each forecast into four values using the k-nearest neighbors algorithm. Further details are reported in Text S1.

3.3 Reservoir performance evaluation

The value of seasonal streamflow forecasts—here measured in terms of hydropower production—certainly depends on predictive skill; however, a second important factor influencing forecast value are the reservoir characteristics. For example, a reservoir constrained by small turbine capacity may perform adequately well utilizing control rules alone as storage is sufficient to buffer inflow variability. Thus, we are not only interested in quantifying forecast value, but also in understanding how value varies as a function of both skill and reservoir characteristics.

3.3.1 Impact of design characteristics for a perfect forecast-based approach

Initially excluding the effect of actual forecast skill, the following performance metric represents the expected improvement from perfect forecast-informed operations as compared to control rules-based operations:

$$I_{PF} = \frac{H_{PF} - H_{ctrl}}{H_{PF}} \times 100\%, \tag{6}$$

where H_{PF} and H_{ctrl} represent the total hydropower production (for the period 1958–2000) obtained with perfect forecast-informed operations and control rules, respectively. A negative value indicates that the control rules outperform the (perfect) forecast-informed operations, whereas a positive value suggests that forecast-informed operations could be beneficial.

To understand how reservoir characteristics may influence benefits based on a perfect forecast approach, we proceed in two steps. First, we label each dam as *success* or *failure* depending on whether the associated value of I_{PF} is larger or smaller than the mean value of I_{PF} across all dams. Note that *failure* implies that the control rules and (perfect) forecast-informed operations generate a similar amount of hydropower, and thus storage and previous month inflow quantities are sufficient for near-optimal release decisions. Second, we explain the likelihood of achieving success through a logistic regression model in which the probability of the binary response variable taking a particular value is a function of the predictor variables. We consider two predictors, namely (1) the ratio of reservoir storage capacity to the mean monthly inflow (x_{fill}, measured in months), and (2) the ratio of maximum reservoir depth to maximum hydraulic head (x_{depth}). The second predictor varies between 0 and 1, and indicates the extent to which hydraulic head is dependent on the depth of the reservoir. The logistic regression model is cross-validated with a 10-fold cross-validation scheme, and evaluated using two metrics, accuracy and Cohen's kappa (McHugh, 2012). Accuracy is the ratio of correctly predicted observations (true positives and true negatives) to the total number of observations. Cohen's kappa is an adjusted accuracy score that accounts for the possibility of correct predictions occurring by chance. The modelling exercise is carried out with the R package *caret*. For additional details, please refer to Text S2, and Table S1–S2 in the Supporting Information.

3.3.2 Impact of forecast skill and design characteristics for a realistic forecast-based approach

Integrating realistic forecasts in lieu of perfect forecast information, we introduce the following performance metric:

$$I_{DF} = \frac{H_{DF} - H_{ctrl}}{H_{PF}} \times 100\%, \tag{7}$$

where H_{DF} represents the total hydropower production (for the period 1958–2000) obtained using realistic forecast-informed operations. I_{DF} is then combined with I_{PF} to calculate the performance metric I that quantifies the potential improvement between realistic and perfect forecast-informed operations:

$$I = \frac{H_{DF} - H_{ctrl}}{H_{PF} - H_{ctrl}} = \frac{I_{DF}}{I_{PF}}. \tag{8}$$

A value of I equal to 1 indicates that benefits from the actual forecasts equal those utilizing perfect forecasts. A value of 0 denotes performance equivalent to applying the control rules only, while a negative value implies that the forecast-informed scheme is inferior to the control rules only approach. We calculate this metric only for the subset of dams achieving a value of I_{PF} greater than the mean value of I_{PF} to better understand if the benefits modeled with perfect forecasts may be attainable with realistic forecasts.

To explain how the metric I varies, we use a linear regression model accounting for both forecast skill and reservoir characteristics. The predictor characterizing the forecast skill is x_{MdAPE}, the median absolute percentage error of the forecast, used in place of KGE because it shows a higher correlation with I. (While KGE gives a broad view of the forecast skill by comparing correlation, mean, and standard deviation of the predicted and observed inflows, $MdAPE$ looks at the forecast error at every time step of the inflow time series. This may make $MdAPE$ a more suitable predictor, since the error at each time step affects the release decisions and, ultimately, the hydropower production.) The second predictor is x_{exceed}, the fraction of time that inflow exceeds the maximum turbine release rate. For more details on the choice of predictors, please refer to Text S3 and Table S3–S4.

4 Results

In this section, we first present the accuracy of the inflow prediction models (Section 4.1) and performance of the forecast-informed schemes (Section 4.2). Then, we quan-

tify the extent to which reservoir design characteristics and forecast skill affect the value of seasonal forecasts (Section 4.3). Lastly, we classify all dams according to their potential to benefit from forecasts, and identify key geographical regions that may benefit the most from forecasts (Section 4.4).

4.1 Potential predictors and accuracy

Reservoir inflow exhibits significant correlation with climate and local drivers (potential predictors) at various monthly lags, and these relationships change across the annual cycle. This has a direct influence on expected predictive skill at each hydropower facility. From a global perspective for monthly dam inflow, the percentage of dams correlated with climate and local drivers in each calendar month varies notably (Figure 2). Evaluating months when a higher percentage of dams are significantly correlated with predictors, some well-known climatic teleconnections can be observed—for example, ENSO and winter-spring streamflow in North America and Europe, NAO and spring-summer peak flows in the northern extratropic regions, and PDO and summer streamflow in southeastern North America and central South America (Lee et al., 2018) (see Figure S1). On average, 27%, 37%, 28%, 20%, and 36% of dams are significantly correlated with ENSO, NAO, PDO, AMO, and snowfall, respectively. Additionally, and not surprisingly, inflow for most dams (72%) exhibits significant 1-month lead autocorrelation. An exception is for some dams during the period March-April, especially in areas with minimal base flow, such as East Asia (Figures 2 and S1). Soil moisture at a 1-month lead presents correlations at 47% of dams across all months with a seasonality similar to the one expressed by the inflow.

Similar to the results of Lee et al. (2018), reservoir inflow and climate predictors are often significantly correlated across several lead months. In these cases, climate predictors are very likely to be included in numerous MP models for various leads, although the correlations may decrease with longer lead-time. When a climate predictor is significantly correlated with reservoir inflow at a 1-month lag (MP1), 74% and 38% of the time that is also included at the 4-month lag (MP4) and 7-month lag (MP7), respectively. Snowfall has a similar retention rate. However, and as expected, autocorrelation in inflow and soil moisture drop more substantially with longer lead; only 53% (28%) of the time when lagged inflow (soil moisture) is included as a predictor in MP1 is it also still included in MP4 (MP7). Globally, an average of 2.7, 1.7, and 0.9 predictors are included in the MP1, MP4, and MP7 models, respectively. In very few cases, the number of predictors increases with longer lead-time. For months when no potential predictors are identified, or either MSESS or GSS is less than zero, the long-term mean inflow for that month is used.

On average, predictors are identified for 8.3 months (MP1), 6 months (MP4), and 4.2 months (MP7) (see Figure 3). As noted previously, a lack of long-lead inflow autocorrelation is predominantly responsible for this drop-off. Prediction accuracy also decreases with lead-time; average KGE values are 0.64 and 0.56 for MP1 and MP7, respectively (Figure 3). Given that prediction accuracy generally declines with lead time, the highest KGE scores across all MP models are associated with MP1 for 68% of dams. For the remaining models, the highest prediction accuracy is recorded for 5% (2%) of dams in the MP4 (MP7) models, emphasizing that skillful climate teleconnections at longer leads do exist, such as in Europe or northwestern and southeastern U.S. (Figure 3b). Globally averaged, each MP model illustrates a significant relationship with KGE scores (r = 0.82–0.96 for all MP models). As for the geographical distribution of KGE, we find relatively high KGE scores in several regions, including North America, eastern South America, Europe, and some regions in western Africa and Asia, where inflows correlate with most of the predictors considered (Figures 3 and S1).

For all MP models, the KGE has an average value of about 0.56, which is regarded as fair skill score. While uniquely tailored forecasts could be produced for each dam considering more local influences, the current prediction approach performs well globally and reflects achievable long-range inflow prediction. Considering the superior performance of the MP1 model, the forecast skill of MP1 only is retained to represent the overall forecast skill for further analyses.

4.2 Performance of forecast-informed operations

The expected performance of perfect and realistic forecast-informed operations is notably different across the 735 hydropower dams (Figure 4). For perfect forecast-informed operations (Figure 4a), substantial increases in hydropower production are possible as compared with the baseline control rules, represented as expected increase in the performance metric I_{PF}. Specifically, 94% of dams exhibit a positive value of I_{PF}; mean improvement is 4.7% and maximum improvement is 60%. For the small number of dams that do not benefit from perfect forecasts, the value of I_{PF} does not drop below -1.7%. Considering all dams collectively, an additional 24 TWh per year of hydroelectricity are generated when adopting the perfect forecast-informed approach in lieu of baseline control rules (IHA, 2019).

When realistic forecast-informed operations are adopted (Figure 4b), a smaller number of dams exhibits increased hydropower production; 25% of dams have a positive value of I_{DF}. These 184 dams show an average improvement of 2.3% and can collectively contribute an additional 1.7 TWh per year in hydropower production. Across all dams, the

maximum and minimum values of I_{DF} are 28% and -24%. This decline in performance, as compared with perfect forecasts, is expected, as realistic forecasts introduce a non-negligible prediction error. Yet, it should also be noted that less than 20% of dams have KGE below 0.5, whereas a disproportionately high number of dams exhibit a negative I_{DF} value—a point on which we shall reurn in Section 4.4. This suggests that for a large number of dams control rule-based operations are superior to realistic forecast-informed operations. For dams with poor I_{DF} and high KGE, the results indicate two points: KGE may not fully capture the relationship between forecast skill and value; and reservoir characteristics may be an important factor influencing the value of realistic forecasts.

4.3 Coincident evaluation of prediction accuracy and reservoir characteristics

To identify the extent to which reservoir characteristics may modulate the value of seasonal forecasts, we identify a logistic regression model that explains the likelihood of achieving success with perfect forecasts (i.e., I_{PF} larger than 4.7%, the mean value of I_{PF} across all dams) as a function of two predictors, x_{fill} (the ratio of reservoir storage capacity to the mean monthly inflow) and x_{depth} (the ratio of maximum reservoir depth to maximum hydraulic head). A 10-fold cross-validation yields a model accuracy and Kappa statistic of 0.785 and 0.535. (Note that the percentage of dams falling into the *success* and *failure* categories is equal to 37% and 63%.)

As illustrated in Figure 5 and Table 1, both predictors influence the probability of achieving success. For x_{fill} values exceeding ten months, dams are highly unlikely to benefit substantially from seasonal forecasts. This suggests that a large storage capacity effectively acts as a buffer against inflow uncertainty. Hence, both control rules and perfect forecast-informed operations tend to attain similar performance. We also observe that some of the smaller dams (x_{fill} <2) fail to attain increased hydropower production even though they are predicted to do so (red triangles in the blue shaded region). This may be attributed to the weekly operations, suggesting that more frequent release decisions may reduce forecast value. For the smaller dams, x_{depth} becomes a critical factor. High values of x_{depth} indicate that the hydraulic head is highly dependent on the reservoir depth, which is in turn dependent on current and near future inflows for dams that cannot accumulate large inflow volumes. Thus, forecast-informed operations become crucial to maintain a high hydraulic head and maximize hydropower production. For hydropower dams that have a low value of x_{depth}, a high hydraulic head is maintained even when storage is low, thereby minimizing the utility of forecasts. These are systems relying on waterfalls, or hilly terrains, to divert part of the water and gain hydraulic head.

Considering only the subset of 269 dams that have an I_{PF} value larger than 4.7%, we apply a linear regression model to estimate the performance metric I. This time, the predictors include x_{MdAPE} (median absolute percentage error of forecast inflows) and x_{exceed} (the fraction of time that inflow exceeds the maximum turbine release rate). The linear regression model has an adjusted R^2 of 0.31—a reasonable performance if we consider that the relationship between forecast skill, value, and reservoir characteristics is explained using two predictors. The reader is referred to Table S3-4 for more complex models that include additional predictors.

The results are presented in Table 2 and illustrated in Figure 6. As expected, higher forecast skill (lower x_{MdAPE}) increases the potential benefits realized by the realistic forecasts; a 1% decrease in x_{MdAPE} increases I by 0.03. Reservoir characteristics can play an important role, as certain configurations allow dams and hydropower production to benefit from realistic forecasts. Specifically, we find that dams that have large fractions of time in which inflow exceeds the maximum turbine release (large values of x_{exceed}) are expected to benefit from forecast-informed operations—even when forecasts are not very accurate (as shown by the diagonal divide in Figure 6). This is predominantly a result of both forecast and observed inflow exceeding the maximum turbine release rate at many time steps, a situation in which the release decision would be the same regardless—inaccurate forecast will not penalize hydropower production.

4.4 A classification of hydropower dams

Banking on the results described above, we divide the dams into four categories on the basis of their potential to benefit from perfect forecast-informed operations (*high potential* if $I_{PF} > 4.7\%$ and *low potential* otherwise) and forecast skill (*good forecast* if $x_{MdAPE} < 20\%$ and *poor forecast* otherwise). The cut-off value for I_{PF} is inherited from the previous analysis (logistic regression model), while the cut-off for x_{MdAPE} divides the 735 dams into two groups containing one third (*good forecast*) and two thirds (*poor forecast*) of the observations. Two groups of dams of particular interest includes (1) dams that fall in regions expressing strong forecast accuracy and have the potential to reap benefits from forecast-informed operations (9% of the total number of reservoirs), and (2) dams with strong potential to benefit from forecast-informed operations but lack forecast accuracy (27%) (Table 3).

As show in Section 4.3, the potential of a dam to benefit from forecasts is largely dependant on its design specifications, which present comparable values in areas with similar orography and design practices. Forecast skill, on the other hand, is largely dependant on climate and teleconnections, which tend to present regional patterns. With

this information at hand, we illustrate in Figure 7 the distribution of the four groups of dams across the thirty climate zones of the Köppen-Geiger climate classification system. Statistics for each climate zone, together with their significance calculated using Chi-squared test, are listed in Table 3. We notice a few interesting regions. First, hydropower dams in Central Europe have neither accurate forecast nor high potential to benefit from forecast-informed operations. This trend is significant ($p < 0.05$) for dams in the Alps (alpine climate, ET), in particular for dams that have large capacity-inflow ratios and high hydraulic heads. Second, hydropower dams in the tropical savanna climate (Aw, Thailand, India, Brazil, and western Africa) and subartic climate (Dfc, Canada, Russia, eastern Europe) have accurate forecast but poor to fair potential. Good forecast skill in the tropical savanna (subartic) climate can be attributed to the high correlation with lagged streamflow, ENSO, PDO, and AMO (snowfall, PDO, and NAO) teleconnections (Figure S1). Third, dams in many regions have high potential to benefit but lack accurate forecast. This is particularly pronounced in the humid subtropical climate (Cwa, Cfa) lying in the southeast regions of Australia, China, U.S., and South America. The trend is also true—but not significant—for dams in Southeast Asia (tropical rainforest, Af) and Pacific Northwest in the U.S. (warm summer Mediterranean, Csb).

5 Discussion

5.1 Implications for planning and management of hydropower projects

In this study, we examine the relationship between seasonal streamflow forecasts and global hydropower production, accounting for the influence of reservoir characteristics. Specifically, we develop seasonal inflow forecasts for 735 headwater dams based on lagged global and local hydro-climatic variables. The forecasts exhibit modest skill globally, but higher skill in several regions, including the snow-dominated northern extratropic region and the tropical savannah climate in mainland Southeast Asia, eastern South America, and western Africa. In agreement with earlier works, our forecasts exhibit well-known teleconnections, such as the one between NAO and spring summer peak flow in northern extratropic regions (Lee et al., 2018) or the one between ENSO and streamflow in Southeast Asia (Sankarasubramanian et al., 2009; Räsänen & Kummu, 2013) and winter-spring streamflow in the Pacific Northwest (Hamlet et al., 2002; Voisin et al., 2006).

We then seek to uncover the relationship between forecast skill, value, and reservoir characteristics by adopting forecasts in the reservoir operations model. While 94% of the dams could benefit from perfect forecasts, only 25% demonstrate improvements when using our realistic forecasts—a fairly low percentage if we consider the forecast skill achieved globally. This highlights the fundamental role of reservoir characteristics in shap-

ing the relationship between forecast skill and value. Key design specifications include a short time-to-fill—a characteristic identified by a few recent studies (Anghileri et al., 2016; Turner, Bennett, et al., 2017; Yang et al., 2020)—an hydraulic head largely dependent on reservoir depth, and inflows often exceeding the maximum turbine release, a design specification that allows operators to work with a larger margin of forecast error during high inflow periods. It is worth stressing here that these results are not intended to provide site operation guidelines, but do represent a first, qualitative, step towards determining the potential benefit of seasonal streamflow forecasts for hydropower operators. The relationships identified here could for example be used to understand how much skill is required for the reservoir inflow forecasts or to characterize the interplay between climatology, hydrology, and dam characteristics in a large region of interest.

By combining information on reservoir characteristics, forecast skill, and climatic zones, we finally identify large regions in which dams would benefit the most from forecasts. A particularly interesting group is formed by dams with strong potential to benefit from forecast-informed operations but lacking adequate forecast accuracy—located in the maritime Southeast Asia, the Pacific Northwest and the humid subtropical climate of the southeast regions of Australia, China, U.S., and South America. These are areas in which watershed-specific analysis may bring immediate benefits to hydropower operators. It is also worth mentioning that the implications of our study go beyond existing reservoirs; dam planning over large scales may also benefit of some of our findings. For example, one could evaluate information on untapped hydropower potential (Y. Zhou et al., 2015; Hoes et al., 2017) and seasonal streamflow predictability to derive some first, qualitative, conclusions on the expected reservoir characteristics. A case in point are run-of-the-river dams: these systems present short time-to-fill and are therefore suitable to implement forecast-informed reservoir sizing and operations (Bertoni et al., 2020).

5.2 Limitations and opportunities

Given the global nature of this study, it is important to note that these results and their implications are not meant to provide site-specific guidelines, but rather qualitative information for watershed-specific studies. Like any other global study, the large spatial domain requires building on a number of assumptions that must be discussed to put our results into perspective.

First, we assume that the goal of dam operators is to maximize hydropower production over the long term (in addition to providing flood control). While this objective provides a tangible indication of forecast value, it may not be fully representative of the local conditions encountered by operators. For example, operators may be inter-

ested to maximize revenue (Anghileri et al., 2018) supply the bulk of power to the grid (Zambon et al., 2012), or complement the generation of other renewable (Graabak et al., 2019). To account explicitly for these aspects, one needs to model the role that dams play in the power market, as recently done for the Western U.S. (Voisin et al., 2020), England (Byers et al., 2020), Laos (A. K. Chowdhury et al., 2020), and the Greater Mekong (K. A. Chowdhury et al., 2020). With these models, one could also infer the willingness to pay for improved streamflow forecasts.

Second, release decisions at individual dams may be affected by joint operations between multiple reservoirs and supported with more accurate data and tailor-made hydrological models than those adopted here. Importantly, these data could include qualitative or quantitative forecasts. Although precipitation and streamflow predictions are not used consistently across the world (Adams & Pagano, 2016), one must acknowledge that medium- to long-range forecasts are increasingly adopted by water utilities—as recently shown by Turner et al. (2020) for 300 dams in the conterminous United States. Access to observed and inferred release decisions could thus help researchers provide a more robust and nuanced estimate of forecast value.

6 Conclusions

Our study suggests that skilful forecasts generally lead to increasing benefits, with such benefits strongly modulated by reservoirs characteristics. However, for some dams, even accurate forecasts may not improve hydropower production if reservoir characteristics are not suitable. On the flipside, some dams can be profitable with little regard to forecast accuracy if their design specifications meet such conditions. Research that integrates these findings with hydrological-electricity models to quantify economic benefits is warranted. Specifically, this may reflect the willingness to pay for improved forecast models. Such an assessment could provide guidance and insights for large-scale hydropower planning and management, particularly as energy systems become more interconnected.

Appendix A Skill scores for model validation

The mean squared error skill score (MSESS) is a deterministic skill score that compares the MSE of prediction model and climatology. It is defined as follows (Wilks, 2011):

$$MSESS = \left(1 - \frac{MSE_{pred}}{MSE_{clim}}\right), \tag{A1}$$

where MSE_{pred} is the MSE of prediction, and MSE_{clim} is the MSE of climatological mean prediction. The perfect score of the MSESS is 1, while a value equal to 0 indicates that the model skill is equal to that of the climatology.

The Gerrity skill score (GSS) is a multi-categorical skill score that rewards correct predictions in rarer categories. The GSS is calculated as follows:

$$GSS = \sum_{i=1}^{3} \sum_{j=1}^{3} p_{ij} s_{ij}, \tag{A2}$$

where p_{ij} is the joint probability of inflow in each category (i, j) of a contingency table (3 x 3 in this study) and s_{ij} is a scoring weight to yield more or less credits based on the frequency of the category (Wilks, 2011). The three categories correspond to the upper, middle, and lower thirds of the inflow during the model calibration period. The GSS ranges from -1 to 1, where a value of 1 represents a perfect forecast and a value of 0 means no predictive skill (compared to the climatology).

Acknowledgments

Jia Yi Ng and Stefano Galelli are supported by Singapore's Ministry of Education (MoE) through the Tier 2 project 'Linking water availability to hydropower supply—an engineering systems approach' (Award No. MOE2017-T2-1-143). All simulations results are available on HydroShare at `http://www.hydroshare.org/resource/ca365ffb1a1f49df8b77e393be965fd8`. We thank the Associate Editor and three anonymous reviewers for their constructive comments.

References

Adams, T. E., & Pagano, T. C. (2016). *Flood forecasting: A global perspective*. Academic Press.

Ahmad, S. K., & Hossain, F. (2019). A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization. *Environmental Modelling & Software*, *119*, 147–165.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., & Siebert, S. (2003). Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences Journal*, *48*(3), 317–337.

Anghileri, D., Castelletti, A., & Burlando, P. (2018). Alpine hydropower in the decline of the nuclear era: Trade-off between revenue and production in the Swiss Alps. *Journal of Water Resources Planning and Management*, *144*(8), 04018037.

669 Anghileri, D., Monhart, S., Zhou, C., Bogner, K., Castelletti, A., Burlando, P., &
670 Zappa, M. (2019). The value of subseasonal hydrometeorological forecasts
671 to hydropower operations: how much does pre-processing matter? *Water*
672 *Resources Research*, *55*, 10159–10178.
673 Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier,
674 D. P. (2016). Value of long-term streamflow forecasts to reservoir opera-
675 tions for water supply in snow-dominated river catchments. *Water Resources*
676 *Research*, *52*(6), 4209–4225.
677 Bertoni, F., Castelletti, A., Giuliani, M., & Read, P. M. (2020). Designing with in-
678 formation feedbacks: Forecast informed reservoir sizing and operation. *Earth*
679 *and Space Science Open Archive ESSOAr*.
680 Bertsekas, D. (1976). *Dynamic programming and stochastic control*. New York, New
681 York: Academic Press.
682 Block, P. (2011). Tailoring seasonal climate forecasts for hydropower operations. *Hy-*
683 *drology and Earth System Sciences*, *15*(4), 1355–1368.
684 Bonnema, M., & Hossain, F. (2017). Inferring reservoir operating patterns across
685 the Mekong Basin using only space observations. *Water Resources Research*,
686 *53*(5), 3791–3810.
687 Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., ...
688 Cottam, A. (2019). A global lake and reservoir volume analysis using a surface
689 water dataset and satellite altimetry. *Hydrology and Earth System Sciences*,
690 *23*(2), 669–690.
691 Byers, E. A., Coxon, G., Freer, J., & Hall, J. W. (2020). Drought and climate
692 change impacts on cooling water shortages and electricity prices in Great
693 Britain. *Nature Communications*, *11*(1), 1–12.
694 Chowdhury, A. K., Dang, T. D., Bagchi, A., & Galelli, S. (2020). Expected bene-
695 fits of Laos' hydropower development curbed by hydroclimatic variability and
696 limited transmission capacity: Opportunities to reform. *Journal of Water*
697 *Resources Planning and Management*, *146*(10), 05020019.
698 Chowdhury, K. A., Dang, T. D., Nguyen, H. T., Koh, R., & Galelli, S. (2020). The
699 Greater Mekong's climate-water-energy nexus: how ENSO-triggered regional
700 droughts affect power supply and CO_2 emissions. *Earth and Space Science*
701 *Open Archive ESSOAr*.
702 De Felice, M., Dubus, L., Suckling, E., & Troccoli, A. (2018, Jul). *The impact of the*
703 *North Atlantic Oscillation on European hydro-power generation.* EarthArXiv.
704 Retrieved from `eartharxiv.org/8sntx` doi: 10.31223/osf.io/8sntx
705 Döll, P., Fiedler, K., & Zhang, J. (2009). Global-scale analysis of river flow alter-

ations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, *13*(12), 2413–2432.

Döll, P., & Lehner, B. (2002). Validation of a new global 30-min drainage direction map. *Journal of Hydrology*, *258*(1-4), 214–231.

Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. (2001). The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. *Geophysical Research Letters*, *28*(10), 2077–2080. Retrieved from `https://doi.org/10.1029/2000GL012745`

Gao, H., Birkett, C., & Lettenmaier, D. P. (2012). Global monitoring of large reservoir storage from satellite remote sensing. *Water Resources Research*, *48*(9).

Gelati, E., Madsen, H., & Rosbjerg, D. (2014). Reservoir operation using El Niño forecasts—case study of Daule Peripa and Baba, Ecuador. *Hydrological Sciences Journal*, *59*(8), 1559–1581.

GEO. (2016). *Global Energy Observatory: Information on Global Energy Systems and Infrastructure.* (Available online at http://globalenergyobservatory.org)

Graabak, I., Korpås, M., Jaehnert, S., & Belsnes, M. (2019). Balancing future variable wind and solar power production in Central-West Europe with Norwegian hydropower. *Energy*, *168*, 870–882.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., ... others (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215. Retrieved from `http://www.nature.com/articles/s41586-019-1111-9`

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, *377*(1-2), 80–91.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., ... others (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, *111*(9), 3251–3256.

Hamlet, A. F., Huppert, D., & Lettenmaier, D. P. (2002). Economic value of long-lead streamflow forecasts for Columbia River hydropower. *Journal of Water Resources Planning and Management*, *128*(2), 91–101.

Hamududu, B., & Killingtveit, A. (2012). Assessing climate change impacts on global hydropower. *Energies*, *5*(2), 305–322.

Hejazi, M. I., Cai, X., & Ruddell, B. L. (2008). The role of hydrologic information in reservoir operation–learning from historical releases. *Advances in Water Resources*, *31*(12), 1636–1650.

Hoes, O. A., Meijer, L. J., Van Der Ent, R. J., & Van De Giesen, N. C. (2017). Systematic high-resolution assessment of global hydropower potential. *PloS one*,

12(2), e0171844.

Hurrell, J. W., & Deser, C. (2010). North Atlantic climate variability: the role of the North Atlantic Oscillation. *Journal of Marine Systems*, *79*(3-4), 231–244. Retrieved from `https://doi.org/10.1016/j.jmarsys.2009.11.002`

ICOLD. (2011). *World Register of Dams. Version Updates 1998–2009* (Tech. Rep.). Paris, France: International Commission on Large Dams. (Available online at www.icold-cigb.net)

IHA. (2019). *2019 hydropower status report* (Tech. Rep.). Retrieved from `https://www.hydropower.org/status2019`

Jolliffe, I. (2002). *Principal component analysis*. Cambridge MA: Springer-Verlag New York. doi: 10.1007/b98835

Kao, S.-C., Sale, M. J., Ashfaq, M., Martinez, R. U., Kaiser, D. P., Wei, Y., & Diffenbaugh, N. S. (2015). Projecting changes in annual hydropower generation using regional runoff data: An assessment of the United States federal hydropower plants. *Energy*, *80*, 239–250.

Kaveh, K., Hosseinjanzadeh, H., & Hosseini, K. (2013). A new equation for calculation of reservoir's area-capacity curves. *KSCE Journal of Civil Engineering*, *17*(5), 1149–1156.

Kim, Y.-O., & Palmer, R. N. (1997). Value of seasonal flow forecasts in Bayesian stochastic programming. *Journal of Water Resources Planning and Management*, *123*(6), 327–335.

Lee, D., Ward, P., & Block, P. (2018). Attribution of large-scale climate patterns to seasonal peak-flow and prospects for prediction globally. *Water Resources Research*, *54*(2), 916–938. doi: 10.1002/2017WR021205

Lehner, B., Czisch, G., & Vassolo, S. (2005). The impact of global change on the hydropower potential of Europe: a model-based analysis. *Energy Policy*, *33*(7), 839–855.

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1), 1–22.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., ... others (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502.

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, *89*(10), 93–94.

Libisch-Lehner, C., Nguyen, H., Taormina, R., Nachtnebel, H., & Galelli, S. (2019).

On the value of ENSO state for urban water supply system operators: Opportunities, trade-offs, and challenges. *Water Resources Research*, *55*(4), 2856–2875.

Liebe, J., Van De Giesen, N., & Andreini, M. (2005). Estimation of small reservoir storage capacities in a semi-arid environment: A case study in the Upper East Region of Ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, *30*(6), 448–454.

Liu, X., Tang, Q., Voisin, N., & Cui, H. (2016). Projected impacts of climate change on hydropower potential in China. *Hydrology and Earth System Sciences*, *20*(8), 3343–3359.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P., & Villars, M. T. (2005). *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: UNESCO.

Maurer, E. P., & Lettenmaier, D. P. (2004). Potential effects of long-lead hydrologic predictability on Missouri River main-stem reservoirs. *Journal of Climate*, *17*(1), 174–186.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276–282.

Ng, J. Y., Turner, S. W., & Galelli, S. (2017). Influence of El Niño Southern Oscillation on global hydropower production. *Environmental Research Letters*, *12*(3), 034010.

Pastor, A., Ludwig, F., Biemans, H., Hoff, H., & Kabat, P. (2014). Accounting for environmental flow requirements in global water assessments. *Hydrology and Earth System Sciences*, *18*(12), 5041–5059.

Payne, J. T., Wood, A. W., Hamlet, A. F., Palmer, R. N., & Lettenmaier, D. P. (2004). Mitigating the effects of climate change on the water resources of the Columbia River basin. *Climatic change*, *62*(1-3), 233–256.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5).

Räsänen, T. A., & Kummu, M. (2013). Spatiotemporal influences of ENSO on precipitation and flood pulse in the Mekong River Basin. *Journal of Hydrology*, *476*, 154–168.

Rheinheimer, D. E., Bales, R. C., Oroza, C. A., Lund, J. R., & Viers, J. H. (2016). Valuing year-to-go hydrologic forecast improvements for a peaking hydropower system in the Sierra Nevada. *Water Resources Research*, *52*(5), 3815–3828.

Sankarasubramanian, A., Lall, U., Devineni, N., & Espinueva, S. (2009). The role of

monthly updated climate forecasts in improving intraseasonal water allocation. *Journal of Applied Meteorology and Climatology*, *48*(7), 1464–1482.

Soncini-Sessa, R., Weber, E., & Castelletti, A. (2007). *Integrated and participatory water resources management–Theory* (Vol. 1). Amsterdam, NL: Elsevier.

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

Turner, S. W., & Galelli, S. (2016). Water supply sensitivity to climate change: an R package for implementing reservoir storage analysis in global and regional impact studies. *Environmental Modelling & Software*, *76*, 13–19.

Turner, S. W., Hejazi, M., Kim, S. H., Clarke, L., & Edmonds, J. (2017). Climate impacts on hydropower and consequences for global electricity supply investment needs. *Energy*, *141*, 2081–2090.

Turner, S. W., Ng, J. Y., & Galelli, S. (2017). Examining global electricity supply vulnerability to climate change using a high-fidelity hydropower dam model. *Science of the Total Environment*, *590*, 663–675.

Turner, S. W., Xu, W., & Voisin, N. (2020). Inferred inflow forecast horizons guiding reservoir release decisions across the united states. *Hydrology and Earth System Sciences*, *24*, 1275–1291.

Van Beek, L., Wada, Y., & Bierkens, M. F. (2011). Global monthly water stress: 1. water balance and water availability. *Water Resources Research*, *47*(7).

Van Vliet, M. T., Wiberg, D., Leduc, S., & Riahi, K. (2016). Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nature Climate Change*, *6*(4), 375.

Voisin, N., Dyreson, A., Fu, T., O'Connell, M., Turner, S. W., Zhou, T., & Macknick, J. (2020). Impact of climate change on water availability and its propagation through the Western US power grid. *Applied Energy*, *276*, 115467.

Voisin, N., Hamlet, A. F., Graham, L. P., Pierce, D. W., Barnett, T. P., & Lettenmaier, D. P. (2006). The role of climate forecasts in Western US power planning. *Journal of applied meteorology and climatology*, *45*(5), 653–673.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D., & Kummu, M. (2014). Annual flood sensitivities to El Niño–Southern Oscillation at the global scale. *Hydrology and Earth System Sciences*, *18*(1), 47–66.

Weedon, G., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., ... Best, M. (2011). Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology*, *12*(5), 823–848.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Yang, G., Guo, S., Liu, P., & Block, P. (2020). Integration and evaluation of forecast-informed multiobjective reservoir operations. *Journal of Water Resources Planning and Management*, *146*(6), 04020038.

You, J.-Y., & Cai, X. (2008). Determining forecast and decision horizons for reservoir operations under hedging policies. *Water resources research*, *44*(11).

Zambon, R. C., Barros, M. T. L., Lopes, J. E. G., Barbosa, P. S. F., Francato, A. L., & Yeh, W. W.-G. (2012). Optimization of large-scale hydrothermal system operation. *Journal of Water Resources Planning and Management*, *138*(2), 135-143.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2015). A global boom in hydropower dam construction. *Aquatic Sciences*, *77*(1), 161–170.

Zeng, R., Cai, X., Ringler, C., & Zhu, T. (2017). Hydropower versus irrigation—an analysis of global patterns. *Environmental Research Letters*, *12*(3), 034006.

Zhang, X., Li, H.-Y., Deng, Z. D., Ringler, C., Gao, Y., Hejazi, M. I., & Leung, L. R. (2018). Impacts of climate change, policy and water-energy-food nexus on hydropower development. *Renewable energy*, *116*, 827–834.

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). ENSO-like interdecadal variability: 1900–93. *Journal of climate*, *10*(5), 1004–1020. Retrieved from `https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2`

Zhao, T., Yang, D., Cai, X., Zhao, J., & Wang, H. (2012). Identifying effective forecast horizon for real-time reservoir operation under a limited inflow forecast. *Water Resources Research*, *48*(1).

Zhou, T., Voisin, N., & Fu, T. (2018). Non-stationary hydropower generation projections constrained by environmental and electricity grid operations over the western United States. *Environmental Research Letters*, *13*(7), 074035.

Zhou, Y., Hejazi, M., Smith, S., Edmonds, J., Li, H., Clarke, L., ... Thomson, A. (2015). A comprehensive view of global potential for hydro-generated electricity. *Energy & Environmental Science*, *8*(9), 2622–2633.

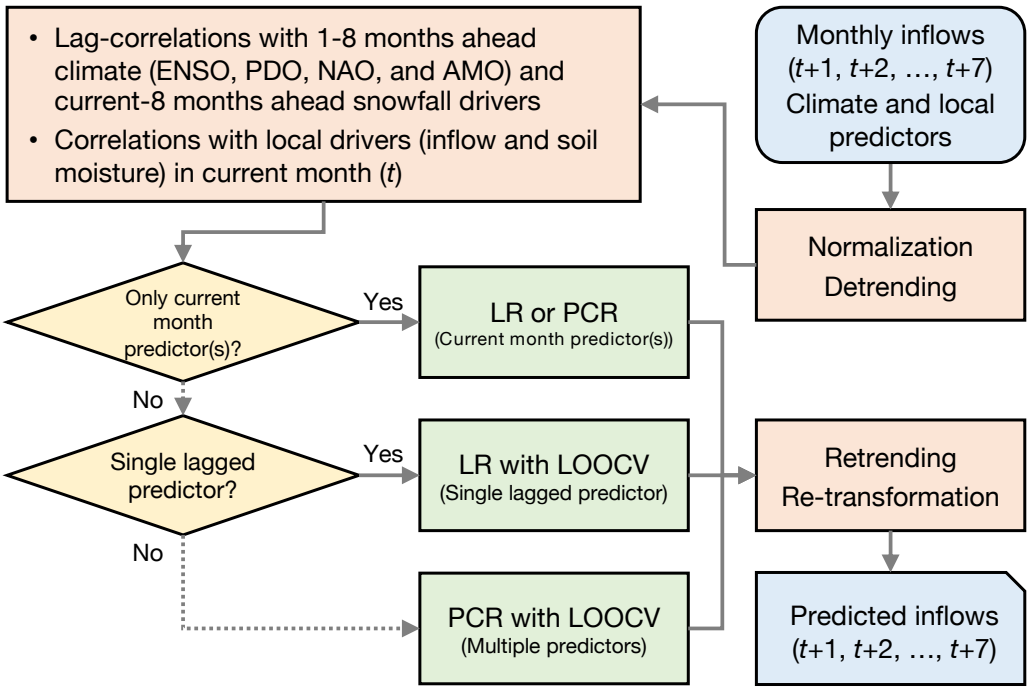

Figure 1. Graphical representation of the monthly prediction (MP) model scheme. At each calendar month t, we develop seven independent models to predict monthly inflows for the next seven months: MP1 $(t+1)$, MP2 $(t+2)$, ..., MP7 $(t+7)$.

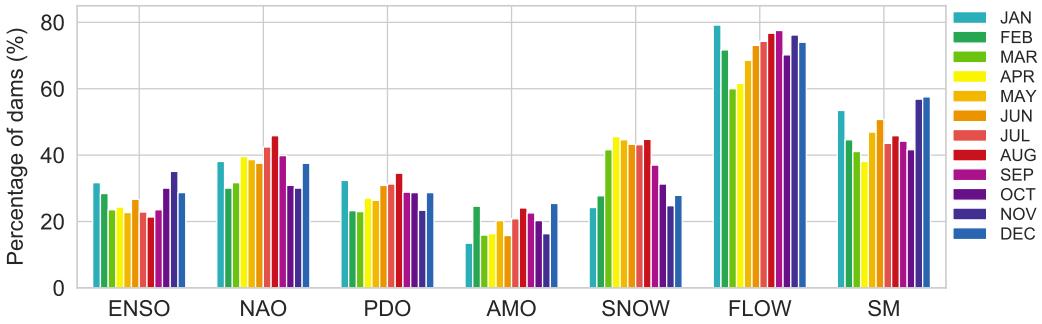

Figure 2. Percentage of dams significantly correlated with lagged predictors (ENSO, NAO, PDO, AMO, and snowfall) and 1-month ahead predictors (inflow and soil moisture) in each calendar month.

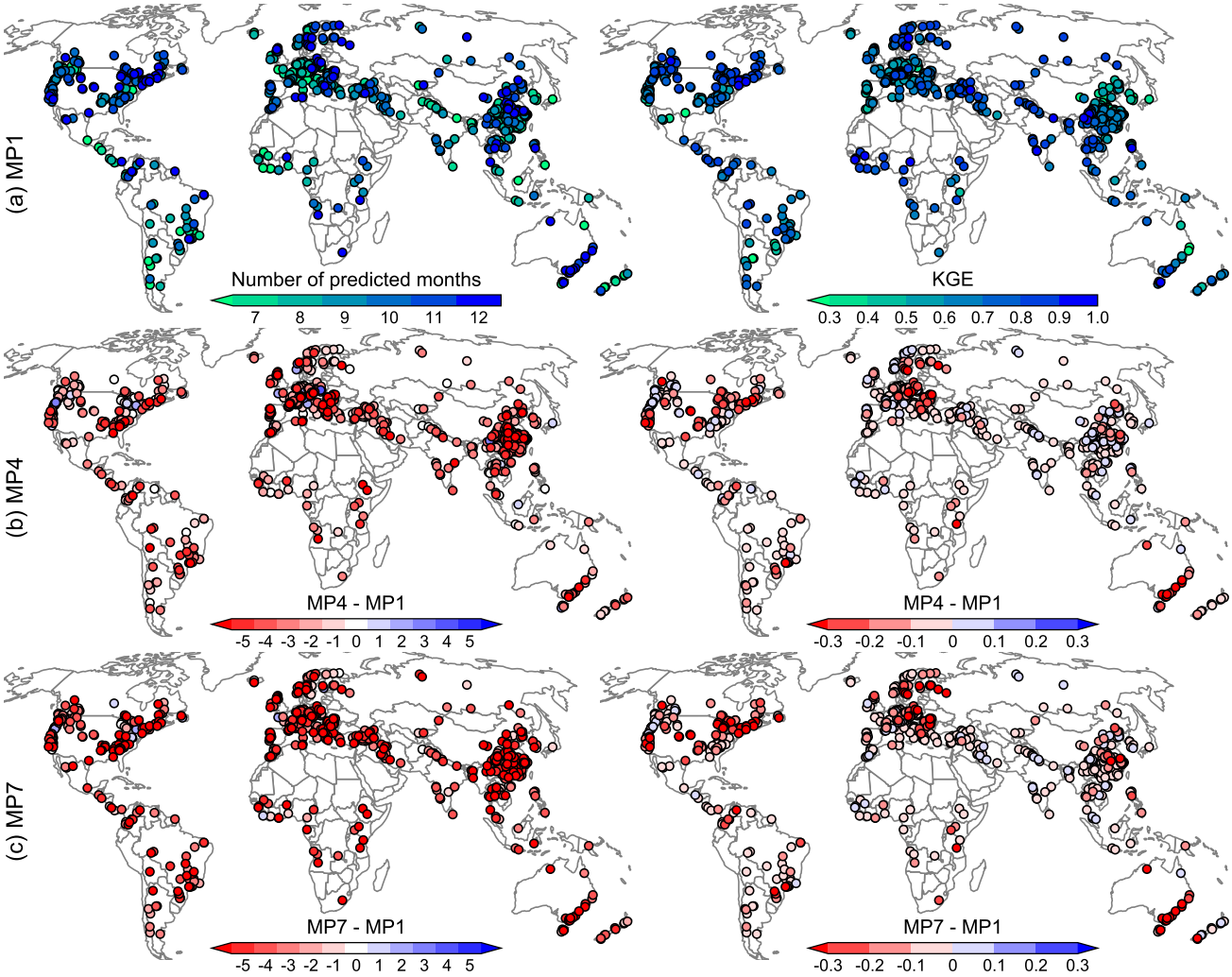

Figure 3. Number of predicted months (left) and KGE (right) of MP1 (a) and differences with respect to MP4 (b) and MP7 (c).

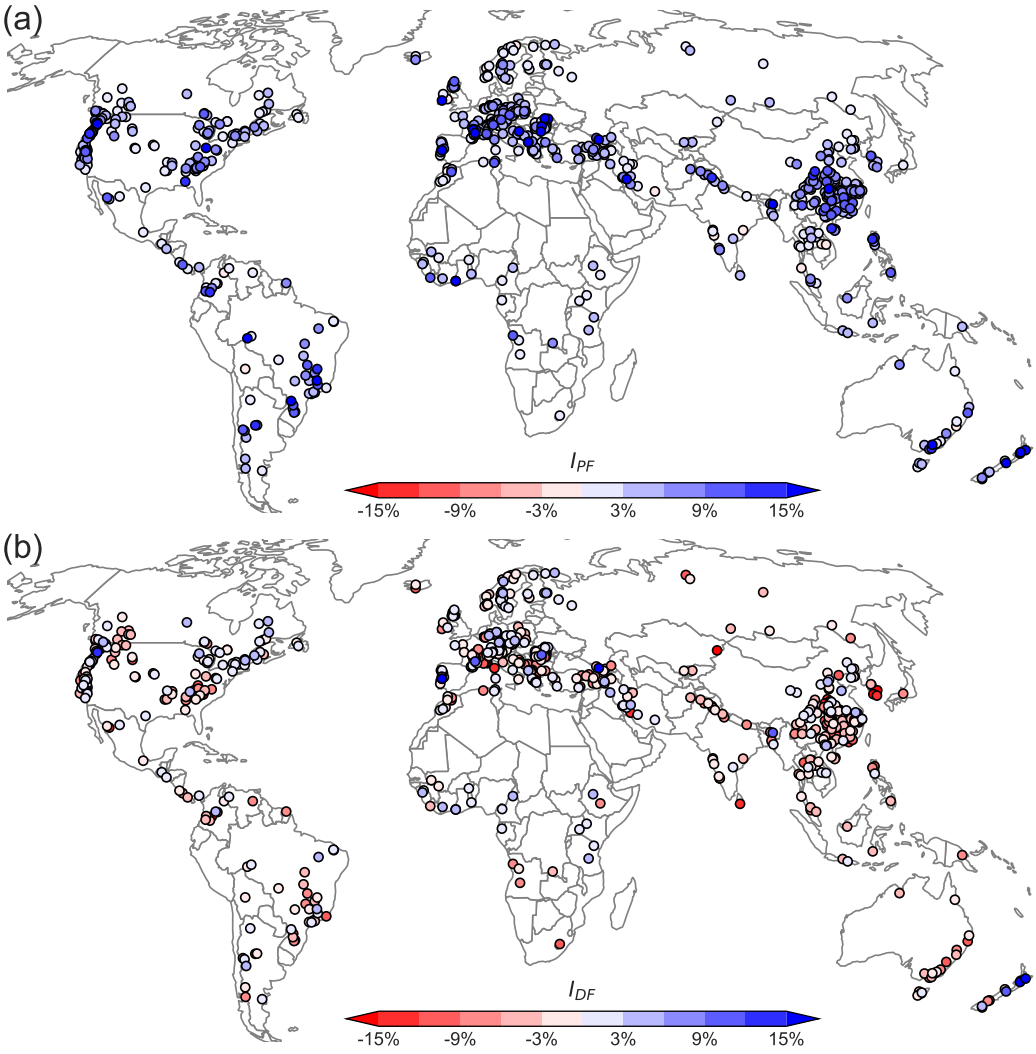

Figure 4. Improvements in hydropower production using perfect (a) and realistic (b) forecasts. The terms I_{PF} and I_{DF} indicate the relative improvement in hydropower production (with respect to the basic control rules) provided by perfect and realistic forecasts. Nearly all dams are able to benefit from perfect forecasts, but only 25% of dams benefits from realistic forecasts.

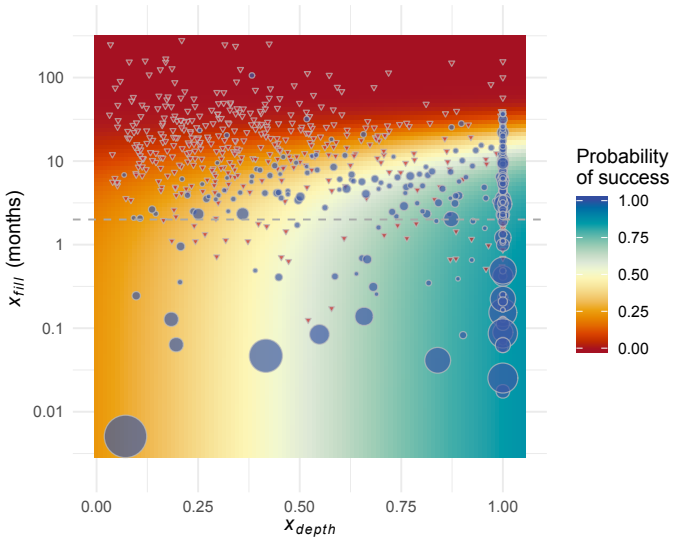

Figure 5. Probability of *success* estimated using a logistic regression model with predictors x_{depth} and x_{fill} (in log scale). Red corresponds to a probability of success equal to zero, meaning that the dam is likely to do well with the control rules. Blue represents a probability of success equal to 1, meaning that a dam is likely to benefit from forecast-informed operations. Each point in the plot represents one of the 735 dams. Blue circles represent dams labelled as *success* ($I_{PF} > 4.7\%$) and red triangles represents *failures*. The size of the blue circles represents the value of I_{PF}. All red triangles have the same size. Dams below the dashed line ($x_{fill} = 2$) are operated with a weekly time step. Dams with low values of x_{fill} (small storage capacity relative to inflow rate) and high x_{depth} (lacking a natural waterfall) are more likely to benefit from forecast-informed operations.

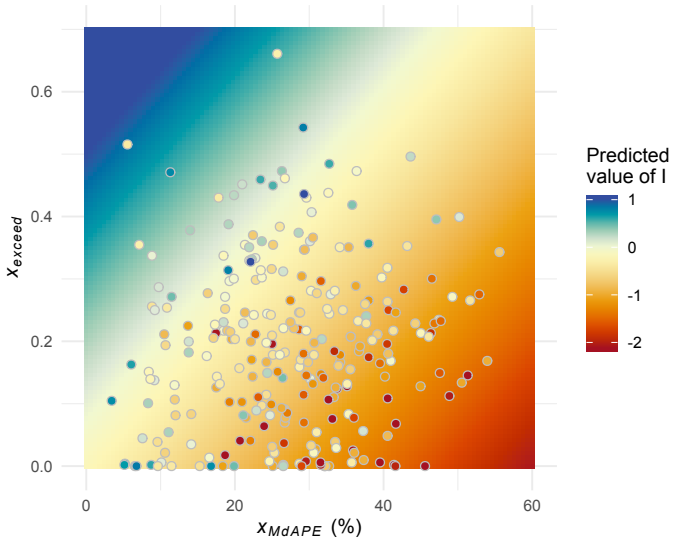

Figure 6. Potential benefits realized by realistic forecast (I) predicted using linear regression with predictors x_{exceed} and the median absolute percentage error (x_{MdAPE}). Red corresponds to negative values of I, meaning that the performance of realistic forecasts is worse that the one attained by control rules. Blue corresponds to positive values of I, meaning that realistic forecasts outperform control rules. Each point corresponds to one of the 269 dams with $I_{PF} > 4.7\%$. The corresponding color represents the value of I attained via simulation with the reservoir operation model. Dams with accurate forecasts and high values of x_{exceed} (inflow frequently exceeds maximum turbine release) tend to have greater hydropower benefits realized from realistic forecasts.

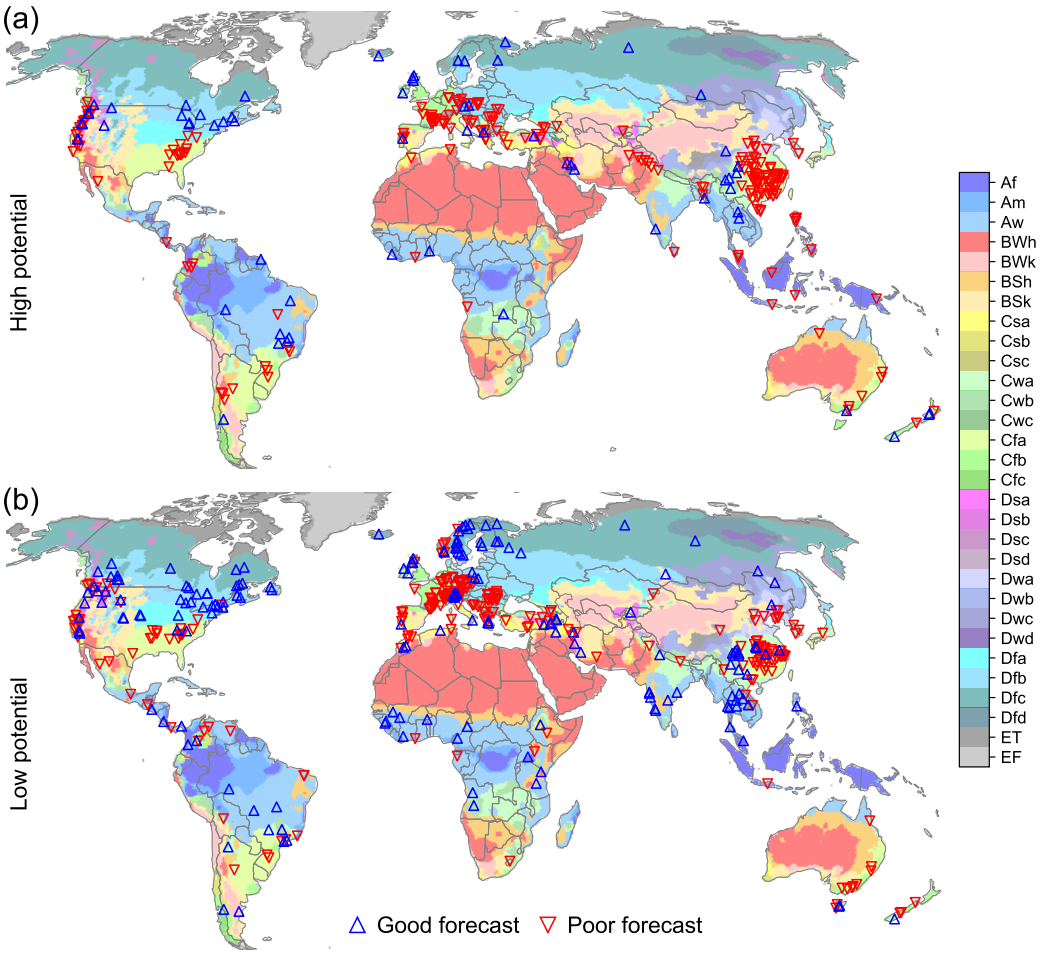

Figure 7. Distribution of dams across climate zones based on their potential to benefit from perfects forecasts. The top (a) and bottom (b) panels represent dams with 'high potential' ($I_{PF} > 4.7\%$) and 'low potential' ($I_{PF} \leq 4.7\%$), respectively, while 'good forecast' and 'poor forecast' represent dams with $MdAPE$ less or greater than 20%, respectively.

Table 1. Coefficients of logistic regression to predict if $I_{PF} > 4.7\%$. The term 'Estimate' represents the increase in log-odds of a dam attaining *success* per unit increase in the value of the predictors.

Predictors	Estimate	Std. Error	Z-value	Pr(>|z|)
(Intercept)	-1.16	0.25	-4.62	<0.01
x_{depth}	2.84	0.31	9.25	<0.01
x_{fill}	-0.10	0.01	-8.76	<0.01

Table 2. Coefficients of linear regression to predict I.

Predictors	Estimate	Std. Error	Z-value	Pr(>|z|)
(Intercept)	-0.18	0.12	-1.485	0.139
x_{MdAPE}	-0.03	0.004	-8.554	<0.01
x_{exceed}	2.36	0.30	7.752	<0.01

Table 3. Distribution of dams across climate zones. H and L represent *high/low potential*, G and P represent *good/poor forecast*. Columns 2-5 (6-9) are the number (percentages) of dams in each group. Column 10 and 11 are the percentages of dams with *high potential* and *good forecast* respectively and are in bold if the observed frequency is different from the expected frequency (global average in final row) significantly ($p < 0.05$ using Chi-squared test).

Climate	HG	HP	LG	LP	HG%	HP%	LG%	LP%	High%	Good%
Af	0	9	3	3	0.00	0.60	0.20	0.20	0.60	0.20
Am	4	6	5	4	0.21	0.32	0.26	0.21	0.53	0.47
Aw	7	6	26	4	0.16	0.14	0.61	0.09	0.30	**0.77**
BWh	2	2	2	0	0.33	0.33	0.33	0.00	0.67	0.67
BWk	0	3	1	0	0.00	0.75	0.25	0.00	0.75	0.25
BSh	1	2	4	6	0.08	0.15	0.31	0.46	0.23	0.39
BSk	0	2	6	5	0.00	0.15	0.46	0.39	0.15	0.46
Csa	4	11	9	19	0.09	0.26	0.21	0.44	0.35	0.30
Csb	6	15	4	10	0.17	0.43	0.11	0.29	**0.60**	0.29
Cwa	5	25	8	16	0.09	0.46	0.15	0.30	**0.56**	0.24
Cwb	1	3	7	2	0.08	0.23	0.54	0.15	0.31	0.62
Cfa	2	56	6	58	0.02	0.46	0.05	0.48	**0.48**	**0.07**
Cfb	9	15	9	41	0.12	0.20	0.12	0.55	0.33	0.24
Dsa	0	2	1	2	0.00	0.40	0.20	0.40	0.40	0.20
Dsb	2	3	5	4	0.14	0.21	0.36	0.29	0.36	0.50
Dsc	0	1	1	1	0.00	0.33	0.33	0.33	0.33	0.33
Dwa	0	5	1	8	0.00	0.36	0.07	0.57	0.36	0.07
Dwb	1	1	2	1	0.20	0.20	0.40	0.20	0.40	0.60
Dwc	2	1	2	0	0.40	0.20	0.40	0.00	0.60	0.80
Dfa	0	1	6	5	0.00	0.08	0.50	0.42	0.08	0.50
Dfb	13	21	32	41	0.12	0.20	0.30	0.38	0.32	0.42
Dfc	7	10	33	20	0.10	0.14	0.47	0.29	**0.24**	**0.57**
ET	1	2	7	36	0.02	0.04	0.15	0.78	**0.07**	**0.17**
Total	67	202	180	286	0.09	0.28	0.25	0.39	0.37	0.34