Received: 13 April 2021 | Revised: 17 December 2021 | Accepted: 5 January 2022

DOI: 10.1002/met.2047

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

RESEARCH ARTICLE

# The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting

**Robert M. Graham**[1] | **Jethro Browell**[1,2] | **Douglas Bertram**[3] | **Christopher J. White**[3]

[1]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, Scotland

[2]School of Mathematics and Statistics, University of Glasgow, Glasgow, Scotland

[3]Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, Scotland

**Correspondence**
Jethro Browell, School of Mathematics and Statistics, University of Glasgow, Glasgow, Scotland.
Email: jethro.browell@glasgow.ac.uk

**Funding information**
This research was funded through the University of Strathclyde's Low Carbon Power and Energy programme, which is a strategic industrial collaboration between the University, Scottish and Southern Energy, Scottish Power and Wood. Jethro Browell is supported by EPSRC Innovation Fellowship (EP/R023484/1). We thank Andrew F. Low and Richard Hearnden from SSE Renewables for their valuable input on this project and providing the historical inflow record for the case study reservoir. The authors declare no conflicts of interest.

## Abstract

Inflow forecasts play an integral role in the management and operations of hydropower reservoirs. In Scotland, the horizon of inflow forecasts is limited in range to approximately 2 weeks ahead. Additional forecast information in the sub-seasonal to seasonal (S2S) range would allow operators to take proactive action to mitigate weather-related risks, thereby improving water management and increasing revenue. The aim of this study is to develop methods of deriving skilful S2S probabilistic inflow forecasts for hydropower reservoirs in Scotland, without the application of a hydrological model. We forecast inflow for a case study reservoir using a linear regression model, trained on historical S2S precipitation predictions and observed inflow rates. Ensemble inflow forecasts generated from the regression model are post-processed using Ensemble Model Output Statistics, to create calibrated S2S probabilistic forecasts. We evaluate forecast skill for 11 different horizons, using inflow observations. Probabilistic forecasts of weekly average inflow rates hold fair skill relative to climatology up to 6 weeks ahead (fCRPSS = 0.01). Forecasts of 28-day average inflow rates hold good skill (fCRPSS = 0.19). The S2S probabilistic inflow forecasts are most skilful during winter, when there is greatest risk of reservoirs spilling. Forecasts struggle to predict high summer inflows even at short lead times. The potential for the S2S probabilistic inflow forecasts to improve water management and deliver increased economic value is explored using a stylized cost model. While applied to hydropower forecasting, the results and methods presented here are relevant to broader fields of water management and S2S forecasting applications.

**KEYWORDS**

ensemble forecasts, hydrology, hydropower, precipitation, S2S, streamflow, sub-seasonal to seasonal prediction, water management

# 1 | INTRODUCTION

Hydropower provides a crucial source of flexibility to many national power systems, and will become increasingly valuable as countries expand the capacity of variable renewable energy resources, such as wind and solar power (Ho et al., 2020; Huertas-Hernando, 2017). Reservoir inflow forecasts play an integral role in the management and operations of hydropower resources. Reliable inflow forecasts help plant operators maximize power generation and prevent spillages, and are used to schedule power generation in advance to coincide with periods of peak energy demand (Ahmad & Hossain, 2019; Huertas-Hernando, 2017; Ødegård et al., 2019). Inflow forecasts are also used to manage water levels and discharge from reservoirs to mitigate downstream flood risks and fulfil environmental requirements.

Hydropower plant operators in Scotland currently use short- and medium-range forecasts that provide sub-daily deterministic inflow predictions up to 14 days ahead. Extending the horizon of inflow forecast could help improve planning and water management decisions, through early and proactive risk mitigation (Ahmad & Hossain, 2019; Klemm & McPherson, 2017; Sene, 2016; Vitart, 2017, 2017). Consider the example of a hydropower plant operator who receives a medium-range inflow forecast showing average inflow rates for the next 7 days. How would the actions of this operator differ if sub-seasonal-to-seasonal (S2S) forecasts were also available, indicating increased probability of above-average inflows for the period from 2 to 6 weeks ahead? The operator may decide to increase power generation over the week ahead. This would lower the reservoir water level and increase capacity for impounding future inflows. Advanced notice of the need to increase power generation reduces the risk of reservoir spills and provides greater flexibility for operators to schedule generation to coincide with periods of peak demand, helping to achieve a higher unit rate for the power generated. However, if the forecast high inflows do not materialize, power may be sold at a lower unit rate in the week ahead, compared with what would have been achieved later in the month. Hence, potential economic benefits from S2S forecast information are highly dependent on the reliability of the forecasts.

The S2S range covers forecast horizons between 2 weeks and 2 months ahead (Vitart, 2014). On S2S timescales, it is not feasible to predict weather conditions for a specific day. However, predictions of average conditions over periods of 7 days or longer are becoming increasingly reliable. Reliable forecast information in the S2S range has been identified as high value to a wide range of industries and users (White, 2017). For example, in addition to hydropower operations, S2S precipitation and streamflow forecasts can play an important role in the management of public water supplies and activation of early warning and response systems for floods and droughts (Arnal et al., 2018; Bell et al., 2017; Svensson, 2015; Vitart, 2017, 2017; White, 2017). S2S forecasts can help end users in many sectors develop proactive management strategies, on weekly timescales, to mitigate weather-related risks. In contrast, with short- and medium-range forecasts, weather events and risks are typically managed with reactive decisions.

The aim of this study is to develop a method of producing skilful S2S probabilistic inflow forecasts for hydropower reservoirs. We focus on a case study reservoir in the Scottish Highlands and evaluate forecast skill for 11 different horizons. We further apply a stylized cost model to ensure the forecasts are sufficiently reliable to improve water management decisions. Finally, we explore how the forecast performance varies for different large-scale weather regimes. While the focus of this paper is on the hydropower sector, the methods and results presented here are relevant to the broader fields of hydrology and S2S prediction.

The paper is organized as follows: Section 2 provides a background information on hydrological forecasting and ensemble weather prediction for S2S timescales; Section 3 documents the datasets used in this study. Section 4 describes the methods used to generate and evaluate the S2S inflow forecasts; in Section 5, we evaluate the skill of the S2S inflow forecasts and explore the potential economic value of these forecasts using a stylized cost model; this is followed by a summary of the main conclusions in Section 6.

# 2 | BACKGROUND

## 2.1 | Hydrological forecasting

Water storage within river catchment areas can act as a long-term memory that provides an important source of skill for predicting future streamflow. Streamflow measures the flow of water in a river channel or stream, whereas inflow measures the flow of water into a reservoir. For river systems with a large water storage capacity, it is possible to produce skilful streamflow forecasts on seasonal or even annual timescales (Bell et al., 2017; Harrigan et al., 2018; Svensson, 2015; Svensson, 2016; Wood & Lettenmaier, 2008). Water storage within a river catchment could include features such as mountain snow cover, aquifers, and soil moisture.

In river systems with a large water storage capacity, future streamflow is largely determined by historical

weather, such as precipitation and evaporation over the previous 3 months. Data required to make streamflow forecasts for these rivers include detailed observations of the present hydrological conditions for the catchment area and/or historical meteorological observations. As an example, recent studies have explored how snow depth measurements could improve inflow forecasts and operations for hydropower reservoirs in Norway (Magnusson et al., 2020; Ødegård et al., 2019). Streamflow forecasts for river systems with large water storage capacities are typically produced using a hydrological model, or assuming that present day streamflow anomalies will persist, or selecting historical streamflow time series that are analogous to the present day conditions (Bell et al., 2017; Harrigan et al., 2018; Magnusson et al., 2020; Svensson, 2015).

Within the United Kingdom, rivers in the southeast of the country typically have the greatest seasonal forecast skill, due to permeable bedrock and aquifers providing a large water storage capacity (Bell et al., 2017; Harrigan et al., 2018; Svensson, 2015; Svensson, 2016). In contrast, in the north and west of the United Kingdom the bedrock is impermeable. In the Scottish Highlands, steep hillsides contribute to fast flowing river systems. Soil water capacity is relatively low, and seasonal snow packs are smaller compared with Scandinavia, the European Alps, or Rocky Mountains. Thus, water storage capacity in these catchment areas is low. Here, skilful streamflow forecasts are dependent foremost on accurate meteorological forecasts (i.e., future weather), rather than detailed observations of past weather and the present day hydrological conditions (Svensson, 2015).

Hydropower operators in Scotland currently use inflow forecasts derived from short- and medium-range deterministic weather predictions. Inflow forecasts are derived from the predicted precipitation falling within the reservoir catchment area. The horizon of deterministic weather forecasts for the mid-latitudes is limited to the medium range, covering periods up to 10–14 days ahead (Zhang et al., 2019). Beyond this range, deterministic forecasts hold no skill due to the chaotic nature of the climate system (Branković et al., 1990; Buizza & Leutbecher, 2015; Zhang et al., 2019).

## 2.2 | S2S forecasting

S2S weather predictions are notoriously challenging, because the timescale is sufficiently long that the climate system's memory of the atmospheric initial conditions is lost, yet too short for ocean variability to play an important role (Vitart, 2004, 2014; Vitart, 2017, 2017). Nonetheless, the skill of S2S predictions has risen steadily over recent years (Vitart, 2014). The majority of the World Meteorological Organisation's Global Producing Centres now issue operational S2S predictions at least once per week (Vitart, 2017, 2017). The improved skill of S2S predictions has been attributed to advances in ensemble prediction systems and the representation of parameterized processes in numerical weather prediction (NWP) models (Vitart, 2014). Ensemble prediction systems produce multiple predictions of the future weather, or ensemble members, which account for uncertainties in the initial atmospheric conditions and model physics. Rather than a single deterministic prediction of the future weather, ensemble prediction systems provide information on the probability of different weather patterns emerging.

Beyond the S2S range is the field of seasonal forecasting or monthly outlooks, which can cover periods up to a year ahead (Johnson, 2019; Merryfield, 2020). Many Global Producing Centres issue seasonal predictions and these are typically once per month (Johnson, 2019), as opposed to S2S predictions that are issued daily or weekly (Vitart, 2017, 2017). The focus of this study is to develop S2S inflow forecasts that can improve water management strategy decisions on weekly timescales.

While increasingly skilful, S2S predictions are imperfect. Systematic biases exist in the output fields and these biases often grow with lead time, meaning that predictions further into the future are associated with larger biases. Post-processing techniques or calibration can reduce systematic biases and significantly enhance forecast skill, compared with raw output fields (Baker et al., 2020; Manrique-Suñén et al., 2020). Forecast calibration requires time series of historical S2S predictions and corresponding observations to identify and correct systematic biases. Hindcast datasets are a vital tool for forecast calibration and evaluation (Vitart, 2017, 2017). These datasets provide historical re-forecasts covering periods of 1–2 decades.

Post-processing techniques for NWP output fields, including S2S predictions, are well established. NWP output fields may be calibrated to model analysis fields using hindcast datasets for general use, or to station data where specific locations are of interest. Many different post-processing methods are available for S2S ensemble predictions (Vannitsem et al., 2018), of which the most widely used approach is called Ensemble Model Output Statistics (EMOS). EMOS typically refers to a form of distribution-based parametric regression, which is a powerful method for post-processing ensemble predictions and creating sharp, calibrated probabilistic forecasts (Gneiting & Katzfuss, 2014).

It is increasingly common to use S2S predictions to create applied forecasts of variables that are impacted by the weather, such as wind, solar, and hydropower output,

or energy demand (Bloomfield et al., 2021; Brayshaw et al., 2020; Lledó et al., 2019; Soret, 2019). Post-processing techniques for these types of forecast are less well established. It is common to calibrate the S2S predictions of the weather variables (e.g., wind speed, precipitation, and temperature) using atmospheric reanalysis datasets (Bloomfield et al., 2021; Lledó et al., 2019; Soret, 2019). These calibrated predictions (e.g., wind speed) are then applied to secondary statistical or physical models to forecast the impact (e.g., wind power output). However, the secondary model may introduce new systematic biases. This means that for optimal performance, the forecasts require further calibration using direct observations of the forecast variable (i.e., wind power). There are two issues here. Often observations of the forecast variable are unavailable. In addition, performing multiple calibration steps runs the risk of compounding errors. If observations of the final forecast variable are available, the simplest and most robust approach is to perform a single calibration step at the end.

## 2.3 | Weather regimes

Variability of weather on sub-seasonal timescales is determined by the manifestation of different weather regimes (Cassou, 2008; Grams et al., 2017; van der Wiel et al., 2019; Vautard, 1990). Weather regimes are formal classifications of large-scale atmospheric circulation patterns that persist for several days or weeks. For the United Kingdom and Europe, different weather regimes fall into two broad categories. The first category is cyclonic regimes. These are characterized by more frequent synoptic frontal systems, and associated with anomalously low mean sea level pressure, greater precipitation, and stronger surface winds. The second category includes blocked regimes, which are associated with higher mean sea level pressure, calm winds, and low precipitation. Blocked regimes typically result in warm temperature anomalies during summer and cold temperature anomalies during winter. Blocked weather regimes have been identified as a key risk to the stability of European power networks during winter, due to a combination of low wind power output and high energy demand for heating that can persist for multiple weeks (Bloomfield et al., 2020; Grams et al., 2017; van der Wiel et al., 2019).

During the winter season, two key weather regimes for the United Kingdom are the positive (NAO+) and negative (NAO−) North Atlantic Oscillation (Bloomfield et al., 2020; Grams et al., 2017; Svensson, 2015; van der Wiel et al., 2019). NAO+ is an example of a cyclonic weather regime, while NAO− is a blocked regime. The relative influence of these two weather regimes can be quantified by the NAO index, which is a standardized measure of the surface pressure difference between the Icelandic low and the Azores high, west of Portugal (Wanner et al., 2001). The phase of the NAO index has a major impact on UK winter precipitation, particularly in western Scotland (Hall & Hanna, 2018; Lavers et al., 2013; Svensson, 2015). NAO+ regimes are associated with significant positive rainfall anomalies, due to more frequent and more intense frontal systems. NAO− regimes are associated with significant rainfall deficits. In contrast with winter, the positive phase of the NAO index during summer is typically associated with blocked weather regimes, bringing high pressure conditions and warmer temperatures. Negative phases of the NAO index in summer are associated with cyclonic weather regimes, bringing lower sea level pressure and large-scale frontal precipitation (Hall & Hanna, 2018).

The skill of S2S predictions is dependent on the ability to predict which weather regimes will emerge over the forecast period. A key source of predictability over the United Kingdom on S2S timescales stems from teleconnections with the Madden–Julian oscillation (MJO) in the tropics (Cassou, 2008; Lee et al., 2019; Vitart, 2017, 2017). Specifically, there is an increased probability of winter NAO+ regimes occurring following an active MJO over the Indian Ocean, while NAO− regimes are more likely to occur following an active MJO over the western Pacific. These teleconnections are further modulated by the El Niño-Southern Oscillation (ENSO) and strength of the polar stratospheric vortex (Lee et al., 2019). These teleconnections between the Tropics and Atlantic are active during the winter season (Lee et al., 2019; Merryfield, 2020; Orsolini et al., 2011; Vitart, 2014; Vitart, 2017, 2017; Vitart & Robertson, 2018; Woolnough, 2019). Hence, the skill of S2S predictions in the North Atlantic region is dependent on these processes being accurately represented in NWP models (Vitart, 2017, 2017).

An alternative approach to using output from NWP for S2S forecasting is to develop statistical forecasts of climate indices for large-scale circulation patterns or weather regimes, such as the NAO index (Cassou, 2008; Hall et al., 2017). Climate index forecasts can be produced using multiple linear regression on statistical predictors, such as MJO and ENSO phases, as well as Atlantic sea surface temperature anomalies and Arctic sea ice concentrations (Cassou, 2008; Hall et al., 2017). Forecasts of climate indices, derived from statistical models or NWP models, can be used to develop applied forecasts based on historical observations assuming there is a strong relationship between the climate index and forecast variable (Hall & Hanna, 2018). Examples of applied forecasts include river flow rates, wind power output, and fault rates on telecommunication networks (Brayshaw et al., 2020; Svensson, 2015). For example,

streamflow, wind power output, and fault rates are all typically higher during winter NAO+ regimes, due to frequent storms bringing strong winds and precipitation to the United Kingdom. The skill of S2S forecasts derived from climate indices depends on two factors: namely, the skill of predicting the climate index, and secondly a strong and accurately modelled relationship between the climate index and forecast variable.

In this study, we develop S2S inflow forecasts for hydropower reservoirs directly from S2S predictions of precipitation. This is a natural extension of the current methods used by hydropower operators, deriving short- and medium-range inflow forecasts from precipitation predictions. The skill of these S2S inflow forecasts will be dependent on the skill of the S2S precipitation predictions and an accurate model to convert precipitation into reservoir inflow.

## 3 | DATA

### 3.1 | Observed reservoir inflow

We generate and evaluate historical S2S inflow forecasts for a large hydropower reservoir in northwest Scotland (Figure 1). The case study reservoir sits at an elevation of 200 m and is surrounded by steep hillsides, without trees. Observations from the reservoir include hourly measurements of water level and power generation. From these data, we calculate the observed reservoir inflow rate. Data from the case study reservoir cover the period from 2009 to 2019.

Observations of the reservoir water level and power generation were first cleaned by removing data points outside the physical limits of these fields or spurious spikes in the time series. Data were flagged where the data value or temporal derivative exceeded threshold values. Approximately 5% of the hourly data points were removed during this cleaning process.

Based on the power generation time series, discharge ($m^3/s$) from the reservoir was calculated as follows:

$$\text{discharge}\,(t) = \frac{\text{power}\,(t)}{\text{efficiency}\,(t) \times \text{density} \times \text{gravity} \times \text{height}\,(t)}$$

where power is the power in Watts (kg $m^2/s^3$) at time $t$; density is the density of water (1000 kg/$m^3$); and gravity is 9.81 m/$s^2$. The efficiency is the turbine efficiency, and height is the net head ($m$). The turbine efficiency and net head vary in time, and are dependent on both the reservoir water level and operating power. Values for the turbine efficiency and net head were estimated at each time interval based on the observed power output and reservoir water level, using efficiency curves provided by the hydropower operator.

In addition to water discharged used to generate power, compensation flows are released from a separate outlet to sustain minimum flow rates in the downstream river

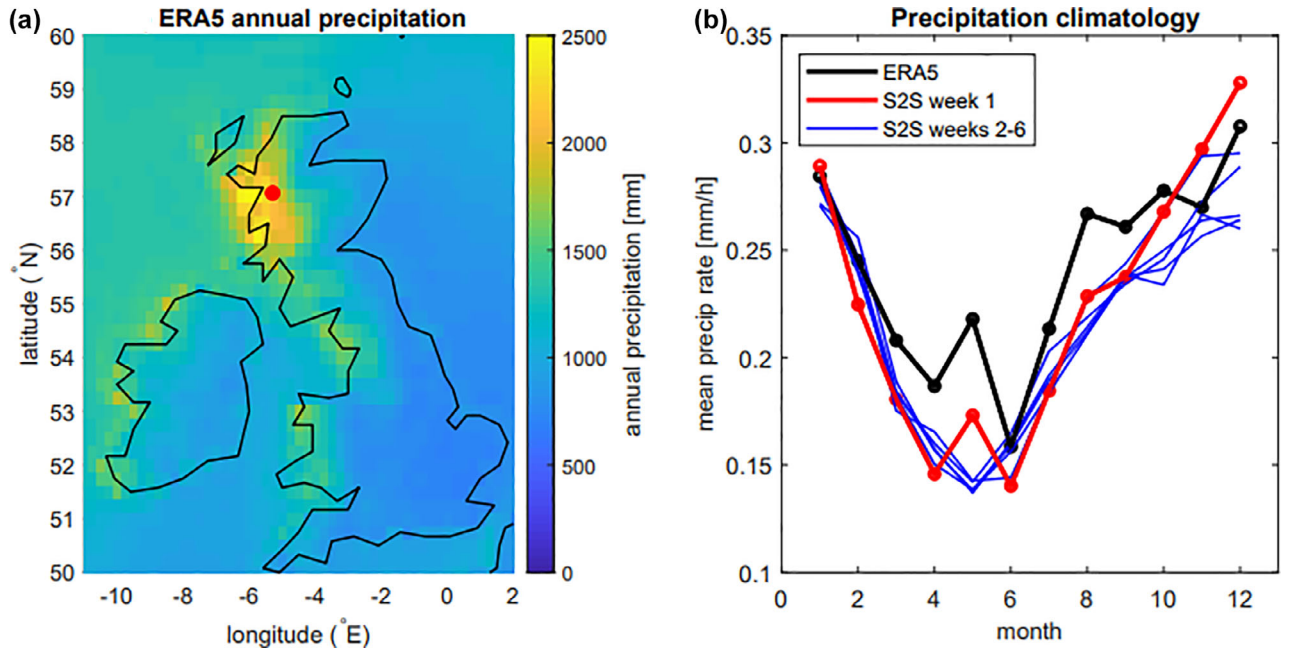


FIGURE 1 (a) ERA5 mean annual precipitation. Red dot indicates location of the case study reservoir. (b) Precipitation climatology for case study reservoir. Lines shows mean monthly rainfall rate for ERA5 (black), week 1 from the S2S predictions (red), and weeks 2–6 of the S2S predictions (blue)

system. These flows are required by legislation to support fish, other wildlife, and vegetation in the catchment.

The volume of water in the reservoir was calculated from the reservoir water level record, using a storage curve equation for the facility. We differentiated the volume of water with respect to time, to calculate the rate of change of water volume per second (units $m^3/s$).

Net inflow ($m^3/s$) was calculated by adding together the discharge from the reservoir, compensation flows, and rate of change of volume:

$$\text{net inflow}\,(t) = \text{discharge}\,(t) + \text{volume change}\,(t) + \text{compensation flow}\,(t)$$

Hourly time series of net inflow contains a lot of noise. However, a clear signal emerges for the daily mean inflow (Figure 2).

In this study, we generate S2S inflow forecasts up to 6 weeks ahead. We evaluate these forecasts over 11 different forecast horizons (Table 1). It is not feasible to reliably predict inflow rates at daily resolution for lead times beyond 2 weeks. Therefore, we evaluate average inflow rates over each forecast horizon, which range in length from 7 days to 42 days (Table 1). S2S inflow forecasts of the average inflow rate over a 42-day period are likely to be valuable for hydropower reservoirs with large storage ratio, defined as the storage volume relative to annual mean inflow (Turner et al., 2020). In contrast, S2S inflow forecasts of 7-day mean inflow rates are more relevant to hydropower reservoirs with a low storage ratio, because a week of high inflow rates could cause a rapid rise in water levels and increase the risk of spilling. In practice, S2S inflow forecasts would be used by hydropower operators alongside existing short- and medium-range inflow forecasts.

For consistency, inflow is always presented as a mean inflow rate averaged over the forecast horizon. To respect anonymity of the site, values presented here are normalized by the annual mean inflow rate for the facility, and thus have no units. Negative values of net inflow exist in the observed inflow record. These indicate a net reduction of water in the reservoir (excluding discharge from the facility), through processes such as evaporation or infiltration combined with low precipitation.

**TABLE 1** Forecast horizons used to evaluate the S2S inflow forecasts and S2S ensemble predictions. We evaluate average inflow rates and average precipitation rates

| Forecast horizon | Forecast period | Forecast length |
|---|---|---|
| Days 1–7 | Week 1 | 7 days |
| Days 8–14 | Week 2 | 7 days |
| Days 15–21 | Week 3 | 7 days |
| Days 22–28 | Week 4 | 7 days |
| Days 29–35 | Week 5 | 7 days |
| Days 36–42 | Week 6 | 7 days |
| Days 1–14 | 2 weeks | 14 days |
| Days 1–21 | 3 weeks | 21 days |
| Days 1–28 | 4 weeks | 28 days |
| Days 1–35 | 5 weeks | 35 days |
| Days 1–42 | 6 weeks | 42 days |

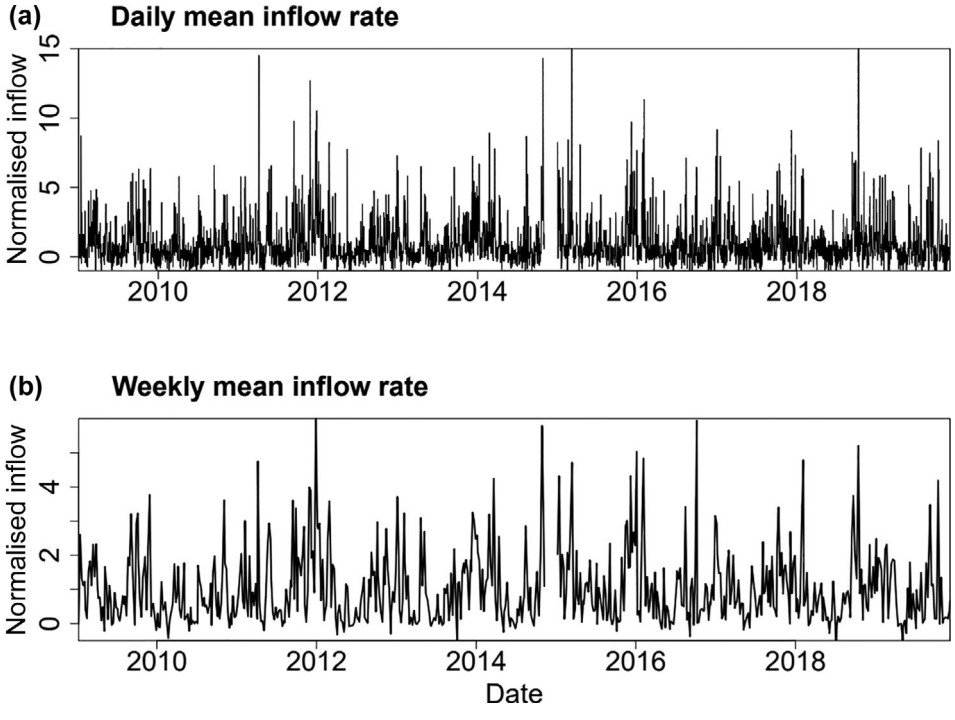


**FIGURE 2** Observed time series of (a) daily and (b) weekly mean net inflow rate for the case study reservoir. Inflow values are normalized (no units) to respect anonymity of the site. Daily and weekly mean inflow rates were divided by the annual mean inflow rate

## 3.2 | S2S ensemble weather predictions

We generate and evaluate historical S2S inflow forecasts using the 2019 hindcast dataset for the European Centre for Medium-range Weather Forecasting (ECMWF) extended-range forecasts (Vitart, 2014). This dataset includes historical S2S weather predictions for the period 1999–2018. These ensemble predictions are issued twice per week, on Mondays and Thursdays, as an extension to medium-range (10-day) predictions. Both are used in this study. The predictions have a forecast horizon of 46 days (6.5 weeks), a temporal resolution of 6 h, and horizontal resolution of 36 km (Vitart, 2014). The historical predictions are comprised of 11 ensemble members, which are produced by perturbing the initial atmospheric conditions and model physics (i.e., the parameters of unresolved processes in the model). The spread of ensemble members provides information on the uncertainty of the predictions and likelihood of different weather patterns emerging.

The hindcast dataset was accessed from the Sub-seasonal-to-Seasonal Prediction Project database (Vitart, 2017, 2017). S2S ensemble prediction datasets from 11 forecasting centres are available through this data portal. All datasets stored in the S2S prediction project archive are interpolated onto a common grid with a horizontal resolution of approximately 150 km. This spatial averaging is justified because the primary source of predictability on S2S timescales originates from large-scale weather patterns. For optimal performance, S2S forecasts require calibration (Manrique-Suñén et al., 2020). This calibration step can be conducted with higher spatial resolution datasets, such as atmospheric reanalyses or observations, which reintroduces signals of local-scale variability. The S2S prediction archive is specifically intended to provide easy access to S2S predictions from multiple agencies for forecast evaluation and calibration studies (Vitart, 2017, 2017) and is widely used (Bloomfield et al., 2021; Brayshaw et al., 2020; Olaniyan et al., 2018; Wang & Robertson, 2019). In this study, the coarser resolution of the S2S Prediction Project dataset (150 km) compared with the original model resolution (36 km) may negatively impact our assessment of skill for the un-calibrated precipitation predictions at the case study reservoir due to smoothing of orographic rainfall patterns in the mountainous terrain. However, using this coarser resolution dataset should have minimal impact on the skill of the final probabilistic inflow forecasts as these are calibrated using in situ inflow observations.

S2S ensemble predictions of large-scale and convective precipitation were extracted from the 150 km grid at the location of the case study reservoir, using bilinear interpolation (Figure 1). We calculate the total precipitation by summing the convective precipitation and large-scale precipitation. For each ensemble member, we average the predicted total precipitation over the 11 different forecast horizons, to give an average precipitation rate (Table 1).

## 3.3 | Atmospheric reanalysis

To evaluate the skill of the initial S2S total precipitation predictions, for the case study reservoir, we use total precipitation from an atmospheric reanalysis. The atmospheric reanalysis used is ERA5, from the ECMWF (Hersbach, 2020, ECMWF, 2017). ERA5 has a horizontal resolution of approximately 30 km, which is comparable to the model resolution of the S2S ensemble predictions (Hersbach, 2020; Vitart, 2014) (Figure 1).

We extract the total precipitation (convective precipitation + large-scale precipitation) from ERA5 at the location of the case study reservoir using bilinear interpolation. These hourly data are averaged over the 11 different forecast horizons (Table 1). We also calculate reanalysis daily mean precipitation rates to explore the relationship between precipitation and observed inflow rates at the case study reservoir.

ERA5 has been shown to capture spatial and temporal patterns of observed precipitation well, even in regions of complex terrain such as the European Alps, North American Rocky Mountains, and Tibetan Plateau (Bandhauer, 2021;Hu & Yuan, 2021; Tarek et al., 2020). EAR5 is compared to ERA-Interim (0.7° resolution) in central Himalaya, where it is found that the coarser resolution reanalysis fails to capture observed precipitation patterns that are present in ERA5 and other higher-resolution reanalysis products (Chen et al., 2021). Biases do exist, and the reanalysis is known to underestimate the intensity and overestimate the duration of high rainfall events (Hu & Yuan, 2021; Tarek et al., 2020). However, the reanalysis has successfully been used in several hydrological modelling applications (Ho et al., 2020; Hu & Yuan, 2021; Tarek et al., 2020). This justifies our use of ERA5 as a reference dataset to evaluate the S2S precipitation predictions, particularly in the absence of an observed precipitation record for the case study reservoir catchment area. ERA5 also covers the entire hindcast period for the S2S predictions. Note that the final S2S inflow forecasts are evaluated using the observed inflow record for the reservoir.

## 3.4 | North Atlantic oscillation index

We explore how the skill of our S2S inflow forecasts varies depending on large-scale circulation patterns. Specifically, we evaluate the skill of forecasts during the extended summer (April–September) and winter (October–March) seasons, as well as for positive and

negative phases of the NAO index. To calculate the NAO index during each forecast horizon, we use a published daily NAO index from the National Weather Service Climate Prediction Centre (Barnston & Livezey, 1987). We average the observed daily NAO index values over each forecast horizon. It is important to note here that this published NAO index is unknown at the time the inflow forecasts are issued. We calculate the forecast skill for each phase of the NAO by considering forecasts for which the observed average NAO index during the forecast period is greater than +0.4 (NAO positive) or less than −0.4 (NAO negative).

# 4 | METHODS

## 4.1 | Deriving inflow forecasts from S2S ensemble weather predictions

We derive inflow forecasts by training a linear regression model for the observed inflow onto the ensemble S2S precipitation predictions. The linear regression model is fitted to a training dataset and verified on "out-of-sample" data. This ensures that error metrics are representative of the true predictive performance and not a result of over-fitting (Messner et al., 2020). We employ a cross-validation scheme to produce and evaluate as many out-of-sample forecasts as possible. Out-of-sample forecasts are produced for each year where inflow observations are available, using a regression model trained on all observations except those from the forecast year and subsequent year. For example, when evaluating inflow forecasts for 2015, inflow observations and S2S ensemble weather predictions from 2015 and 2016 are excluded from the training dataset. This system avoids possible artificial enhancement of the forecast skill through long-term memory in the climate system.

The linear regression model converts a predicted precipitation rate into an inflow rate forecast. The same linear regression model is applied to all horizons of the S2S inflow forecasts (Table 1). This model is trained on the average precipitation rates from days 1–7 of all ensemble S2S predictions in the training dataset and corresponding 7-day mean inflow rate observations. This is a departure from common methods for short- and medium-range forecasts, where typically a separate model is used for each forecast horizon or lead time, in order to account for possible model drift (i.e., biases in the precipitation predictions that grow with lead time). However, for lead times beyond 14 days, there is large uncertainty in ensemble S2S predictions and this is reflected by a large spread among the ensemble members. If we train a linear regression model using precipitation rates from horizons other than days 1–7 of the ensemble S2S predictions, the relationship between predicted precipitation rate and observed inflow rate is weak. This means that the regression model forecasts inflow rates close to climatology for all predicted precipitation rates, resulting in a climatological forecast with an unrealistically narrow spread among ensemble members.

We generate S2S forecasts for the average inflow rate at 11 horizons, which range in length from 1 to 6 weeks (Table 1). For each ensemble member of the S2S predictions, we average the total precipitation rate over the horizon of the inflow forecast. The predicted precipitation rate from this ensemble member is then applied to the linear regression model to calculate an inflow rate forecast. Through this method, we use the linear regression model to generate an ensemble of S2S inflow forecasts for each of the 11 horizons. We refer to these as the benchmark ensemble inflow forecasts.

The benchmark ensemble inflow forecasts are un-calibrated. It is possible to produce more reliable and accurate forecasts by applying post-processing techniques to calibrate the forecasts, correcting any systematic biases (Manrique-Suñén et al., 2020). Here, we apply EMOS post-processing techniques that build on the methods outlined by Scheuerer (2014) for calibrating short-range precipitation forecasts over Germany. In addition to calibrating the forecasts, this post-processing step generates a complete probabilistic forecast distribution, as opposed to an ensemble forecast.

The EMOS model we apply is an example of a generalized additive model for location, scale, and shape (GAMLSS), which is an extremely flexible modelling framework in which the parameters of a given parametric distribution are modelled as linear (generalized) additive models (Rigby & Stasinopoulos, 2005). The probability distribution for each S2S inflow forecast is modelled as a zero adjusted gamma distribution. The choice of distribution and additive model structure was guided by expert judgement and cross-validation on training data.

A separate EMOS model is trained for each of the 11 horizons that we generate S2S inflow forecasts for. This helps account for evolving properties of the ensemble S2S predictions with increasing lead time and model drift, and ensures that the final S2S inflow forecasts for each horizon are calibrated.

The parameters of the zero adjusted gamma distribution for each inflow forecast are modelled as functions of three features derived from the ensemble members of the benchmark inflow forecast (Scheuerer, 2014), in addition to a smoothed function for the day of the year that helps better represent seasonal trends. The three features from the benchmark inflow forecasts include: (1) the ensemble

mean inflow rate; (2) the fraction of ensemble members with a forecast inflow rate less than or equal to zero; and (3) the ensemble mean difference (Scheuerer, 2014). The ensemble mean difference measures that mean absolute difference between each of the ensemble members of any given forecast. Thus, these three features derived from the $K$ ensemble members $m_{1,t},...,m_{K,t}$ at time $t$ of the benchmark inflow forecast are defined as, the ensemble mean inflow

$$\overline{m}_t = \frac{1}{K}\sum_{k=1}^{K} m_{k,t} \tag{1}$$

the fraction of ensemble members less than or equal to zero

$$\overline{1_{\{m=0\}}}_t = \frac{1}{K}\sum_{k=1}^{K} 1_{\{m_{k,t}<=0\}}, \tag{2}$$

and the ensemble mean difference (Scheuerer, 2014)

$$\overline{\Delta m}_t = \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\left|m_{k,t} - m_{k',t}\right| \tag{3}$$

The ensemble mean difference is preferable to other measures of dispersion, such as standard deviation and interquartile range, because it is based on absolute rather than squared differences and is sensitive to all ensemble members (Scheuerer, 2014).

Zero adjusted gamma distributions do not allow negative values. However, in practice, negative values of inflow exist due to the effects of evaporation and infiltration during prolonged periods without precipitation. We therefore add an offset of all observed inflow values, which corresponds to the largest negative inflow rate within each training period. This offset is then subtracted from the modelled probability distributions. The final S2S probabilistic inflow forecasts, following the EMOS post-processing, provide the probability of the average inflow rate over the forecast horizon falling above or below any given threshold.

The zero adjusted gamma distribution $f_{\mathrm{ZAGA}}(\cdot)$ of inflow $y$ is effectively a mixture of a conventional gamma distribution and a probability mass at zero. It is given by

$$f_{\mathrm{ZAGA}}(y|\mu,\sigma,\nu) = \nu\delta(y) + (1-\nu)\left[\frac{1}{(\sigma^2\mu)^{1/\sigma^2}}\frac{y^{\frac{1}{\sigma^2}-1}e^{-y/(\sigma^2\mu)}}{\Gamma(1/\sigma^2)}\right] \tag{4}$$

where $\mu$ and $\sigma$ are the shape and scale parameters of the conventional gamma distribution, respectively, and $\nu$ is the probability of $y=0$. $\delta(\cdot)$ is the Dirac delta function, which is equal to 1 when $y=0$ and zero otherwise, and $\Gamma(\cdot)$ is the gamma function. We model the parameters of the zero adjusted gamma distribution as follows:

$$\log(\mu_t) = \beta_{1,0} + \beta_{1,1}\overline{m}_t + \beta_{1,2}\overline{\mathbf{1}_{\{m=0\}}}_t + s_1(t) \tag{5}$$

$$\log(\sigma_t) = \beta_{2,0} + \beta_{2,1}\overline{m}_t + \beta_{2,2}\overline{\Delta m}_t + s_2(t) \tag{6}$$

$$\mathrm{logit}(\nu_t) = \beta_{3,0} + \beta_{2,1}\overline{m}_t \tag{7}$$

where $s_i(t)$ are cyclic cubic regression splines with a period of 1 year that model smooth seasonal variation, and $\beta_{i,j}$ are parameters to be estimated. This post-processing was performed using the *gamlss* package in R (Rigby & Stasinopoulos, 2005).

## 4.2 | Evaluating the skill of S2S inflow forecasts and ensemble weather predictions

We evaluate the skill of the ensemble S2S precipitation predictions and S2S inflow forecasts for the 11 different horizons using the fair continuous ranked probability skill score (fCRPSS) (Ferro, 2014; Manrique-Suñén et al., 2020). This is a common method used for evaluating ensemble and probabilistic forecasts (Buizza & Leutbecher, 2015; Harrigan et al., 2018; Scheuerer, 2014). The continuous ranked probability score is similar to the mean absolute error for deterministic forecasts and measures the difference between the forecast and observed cumulative density functions. The CRPSS is a measure of the average continuous ranked probability score over the forecast evaluation period, benchmarked against the use of climatological forecasts. The fair version of this skill score (fCRPSS) is designed to reward ensembles that behave as though they are sampled from the same distribution as the verifying observations, which is particularly important for small ensemble sizes (Ferro, 2014; Manrique-Suñén et al., 2020).

An fCRPSS of 1 indicates a perfect forecast, while a score of 0 or below indicates no skill relative to the use of climatological forecasts. Between these extremes, we classify skill scores less than 0.15 as fair, between 0.15 and 0.30 as good, and greater than 0.30 as very good. This follows the convention used by the S2S4E project (https://s2s4e.eu/). These thresholds are somewhat arbitrary as the value of forecasts will depend on the use-case under consideration, which motivates the proceeding section of this paper.

The fCRPSS measures the forecast skill relative to the use of climatological forecasts. Climatological inflow forecasts are based on historical inflow observations from

the case study reservoir. Climatological precipitation forecasts are based on ERA5 precipitation records. Specifically, we average the historical inflow and precipitation rates over the duration of each forecast horizon, and use the historical climatological distribution for the forecast month. Consistent with the method used for splitting observations into training and evaluation datasets, we exclude the observations from the forecast year and subsequent year from the climatological datasets. For example, climatological inflow forecasts for January 2015 would include all historic inflow observations from January, excluding years 2015 and 2016.

### 4.3 | Evaluating economic value of inflow forecasts

In addition to evaluating the statistical skill of the S2S probabilistic inflow forecasts, we assess the potential economic value of these forecasts through improved water management. To measure forecast value, we develop a stylized cost model based on the classical *"News Vendor"* optimization problem (Khouja, 1999). The cost model is guided by current operational practice, with the underlying principle of maintaining the reservoir at a target water level for the time of year.

In the stylized cost model framework, costs fall into two categories. The first category includes costs incurred from precautionary actions taken to maintain the reservoir at its target water level, based on forecast information. The second category includes costs incurred due to the actual observed inflow, following any precautionary actions taken. An example of category 1 costs would be if an inflow forecast indicates an increased probability of high inflows. Based on this forecast information, operators decide to extend the planned generation schedule by 5 h. This increases discharge from the reservoir, in order to accommodate the higher inflows that are forecast. However, the additional power generation would need to take place during periods of lower demand (off-peak energy prices), requiring power to be sold at a lower unit rate, and therefore represent an opportunity cost. Category 2 costs relate to any deviations from the target water level at the end of each forecast period. For example, if the actual observed inflow is greater than the discharge from the reservoir, despite any precautionary adjustments to the generation schedule, the water level in the reservoir will rise. This reduces the potential for impounding future inflows. To prevent future spillages, operators may decide to increase discharge from the reservoir. This also requires extending power generation into off-peak periods, resulting in a lower unit rate of power sold and thus an opportunity cost.

The optimal (risk-neutral) water management decision for each forecast period is to take the precautionary action (i.e., adjust the power generation schedule) that minimizes the expected costs based on the information provided by the inflow forecasts. If the inflow forecasts are reliable and skilful, this will deliver reduced costs relative to using climatological forecasts. The cost model is applied independently to each inflow forecast in the case study dataset (i.e., the 11 forecast horizons, with two forecasts per week from 2009 to 2018). The relative economic value of three types of forecasts (probabilistic, point [p50], and climatological) are evaluated by comparing the total costs incurred for all forecasts over the complete time series. The probabilistic forecasts are the S2S calibrated probabilistic forecasts. Point forecasts correspond to the median (p50) of the predictive distribution of the S2S probabilistic inflow forecasts. Reducing ensemble or probabilistic forecasts down to a single point value (e.g., the ensemble mean) is a common approach, to make these forecasts easier to interpret. Climatological forecasts are also point forecasts, which follow current operational practice, and are defined here as the median (p50) of the historical inflow observations.

In this model framework, the economic value of the different forecasts is sensitive to the choice of peak and off-peak energy price. To explore this dependency, the cost model is run for a range of price differentials (peak–off-peak) from £5/MWh to £100/MWh, while the peak price is kept constant at £50/MWh. The same differential may also result from a higher peak price and non-negative off-peak price. Therefore, the relative change in water value with price differential is more relevant than the absolute value. To test the significance of the results, we apply a bootstrap method to resample the 10-year time series 1000 times (Messner et al., 2020). All results shown indicate the spread of two standard errors based on bootstrap resampling, which approximately represents the 90% confidence interval. For a complete description of the cost model set-up, we refer to the Data S1.

## 5 | RESULTS

### 5.1 | Relationship between precipitation and observed inflow

To explore the relationship between precipitation and inflow at the case study reservoir, we calculate the cross correlation between the ERA5 daily mean precipitation rate and observed inflow rate. This gives a peak correlation coefficient of 0.76 with a 0-day lag-time (Figure 3a). The cross correlations are asymmetric, with higher correlation coefficients for negative lag times (Figure 3). This

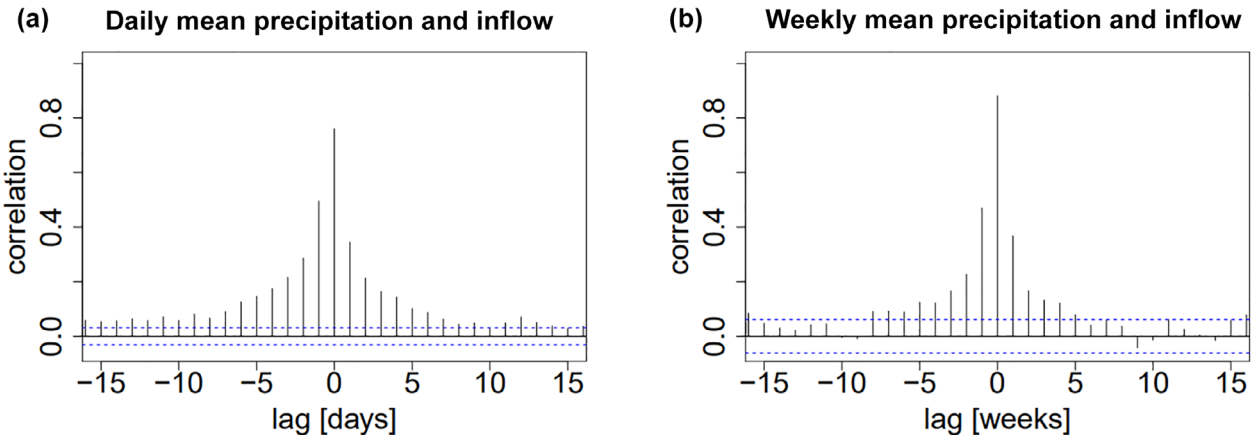


**FIGURE 3** Cross correlation of observed inflow and ERA5 precipitation at the case study reservoir. (a) Daily mean, and (b) weekly mean precipitation and inflow rates. Negative lags indicate how past precipitation is correlated with present inflow

demonstrates that the observed inflow is more strongly correlated with past precipitation, as opposed to future, which is to be expected. The 0-day lag indicates that most precipitation within the catchment area takes less than a day to enter the case study reservoir, confirming that the reservoir is located in a highly responsive catchment area. This is consistent with the steep terrain surrounding the reservoir and relatively low soil water capacity in western Scotland (Bell et al., 2017; Harrigan et al., 2018; Svensson, 2015; Svensson, 2016).

Correlations between the ERA5 precipitation and observed inflow rates are stronger when using weekly average values, giving a peak correlation coefficient of 0.90 (Figure 3b). Several physical processes can explain the higher correlation coefficients found for longer averaging periods, including delays between precipitation and inflow due to freezing temperatures (e.g., snowfall and frozen soil water) and changing soil water content. In addition, temporal averaging smooths random errors in the reanalysis precipitation record. In the remainder of this study, we use predicted average precipitation rates over periods of 7 days or longer to forecast average inflow rates.

The high correlation coefficients found here for the ERA5 precipitation rate and observed inflow demonstrate a strong physical coupling between precipitation and inflow at the case study reservoir (i.e., relationship between the weather variable and impact). These results also indicate that ERA5 provides a reliable precipitation record for the case study site.

## 5.2 | Skill of S2S precipitation predictions

We calculate a skill score of fCRPSS = 0.45 for the average precipitation rate in week 1 (days 1–7) of the S2S predictions at the case study reservoir, relative to the ERA5 climatology (Figure 4a). This demonstrates very good skill. However, the skill of the ensemble predictions decreases rapidly with increasing lead time. We find an fCRPSS of 0.09 for the average precipitation rate in week 2 (days 8–14), reflecting fair skill (Figure 4a). Moving beyond week 2 and into the S2S range, skill scores are negative (Figure 4a). This means that the ERA5 climatological precipitation distributions are more reliable than the S2S precipitation predictions. Note these evaluations of forecast skill are based on un-calibrated S2S predictions, using a coarser horizontal resolution (150 km), compared with the original model output (36 km).

The S2S precipitation predictions follow a similar seasonal cycle for the case study reservoir to the ERA5 precipitation record (Figure 1b). However, during most months, the S2S predictions underestimate precipitation at this location. This may be due to coarse resolution analysed here, resulting in spatial averaging of topographically enhanced rainfall. Both the S2S predictions and ERA5 reveal a minimum in precipitation during spring and early summer, from March to June. Spring precipitation is enhanced during May in ERA5 and week 1 of the S2S predictions. However, this is not replicated for longer lead times in the S2S predictions (Figure 1). This feature is likely due to localized convective rainfall (Lavers et al., 2013). The most significant model drift and spread in precipitation predictions for different lead times is during winter months from October to December (Figure 1).

The timing of precipitation events is a major source of uncertainty in S2S predictions. When considering predictions of the average precipitation rate over periods of time longer than 1 week, this source of uncertainty, meaning there is greater skill in the predictions (Figure 4). Some users may be more interested in

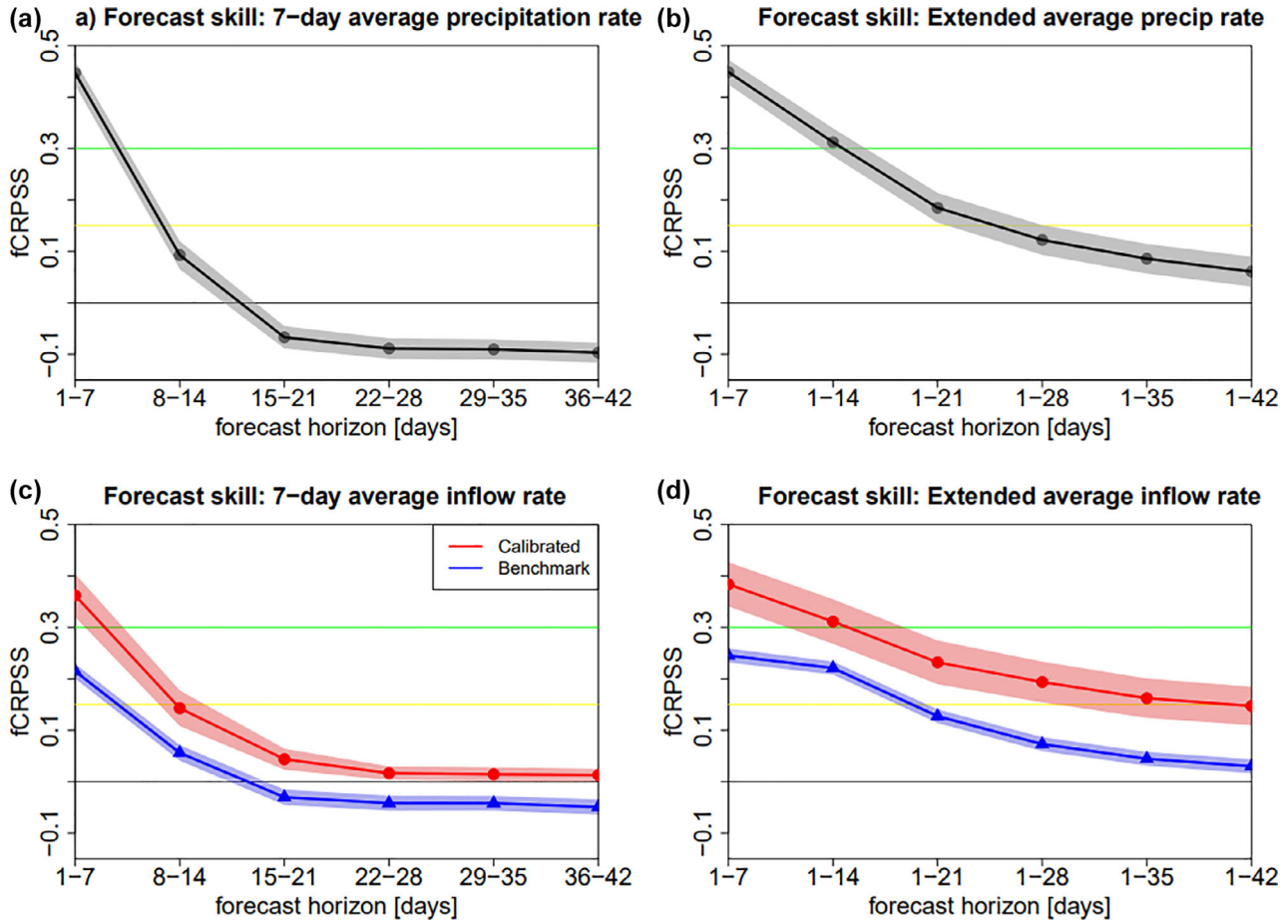


**FIGURE 4** Skill of the precipitation and inflow forecasts at the case study reservoir, evaluated for different forecast horizons (Table 1). (a) Skill of the S2S ensemble weather predictions for 7-day average precipitation rates of different lead times. (b) Skill of the S2S ensemble weather predictions for extended average precipitation rates from 1 to 6 weeks ahead. (c) Skill of the S2S benchmark ensemble (blue) and EMOS probabilistic (red) inflow forecasts for 7-day average inflow rates of different lead times, corresponding to forecast weeks 1–6. (b) Skill of the S2S inflow forecasts for extended average inflow rates from 1 to 6 weeks ahead. Forecast skill is measured using the fair continuous ranked probability skill score, using the SpecsVerification package in R. shaded areas indicate +/−2 standard deviations. Skill of the precipitation predictions is evaluated over the 1999–2018 hindcast period using ERA5 precipitation as the reference dataset. Inflow forecast skill is evaluated over the observation record from 2009 to 2018, using inflow observations as the reference dataset. An fCRPSS value >0 is classed as fair forecast skill, progressing to good forecast skill for values above 0.15 (yellow line) and very good for values above 0.3 (green line)

predictions of total precipitation over an entire month, rather than the total precipitation within each week of that month. Significant greater skill is available to users of long-range average precipitation predictions. For example, predictions of the average precipitation over a 4-week period (days 1–28) hold comparable skill to forecasts for the 7-day average precipitation rate during week 2 (days 8–14) (Figure 4). Fair skill (fCRPSS = 0.06) even exists for predictions of the average precipitation rate over the coming 6 weeks (days 1–42) (Figure 4). It is important to note that the primary source of skill for these extended average precipitation predictions originates from weeks 1 and 2 of the S2S predictions. The skill and reliability of precipitation totals for weeks 1–2 of the S2S predictions provide a strong indication whether the forecast horizon as whole will be wetter or drier than average.

## 5.3 | Skill of S2S inflow forecasts

Consistent with the S2S precipitation predictions, we identify skill for the 7-day average benchmark ensemble inflow forecasts for the horizons of week 1 and week 2, but no further (Figure 4c). The benchmark inflow forecasts hold good skill (fCRPSS = 0.22) for week 1 (days 1–7), and fair skill for week 2 (Figure 4c). For longer averaging periods, the benchmark S2S ensemble inflow

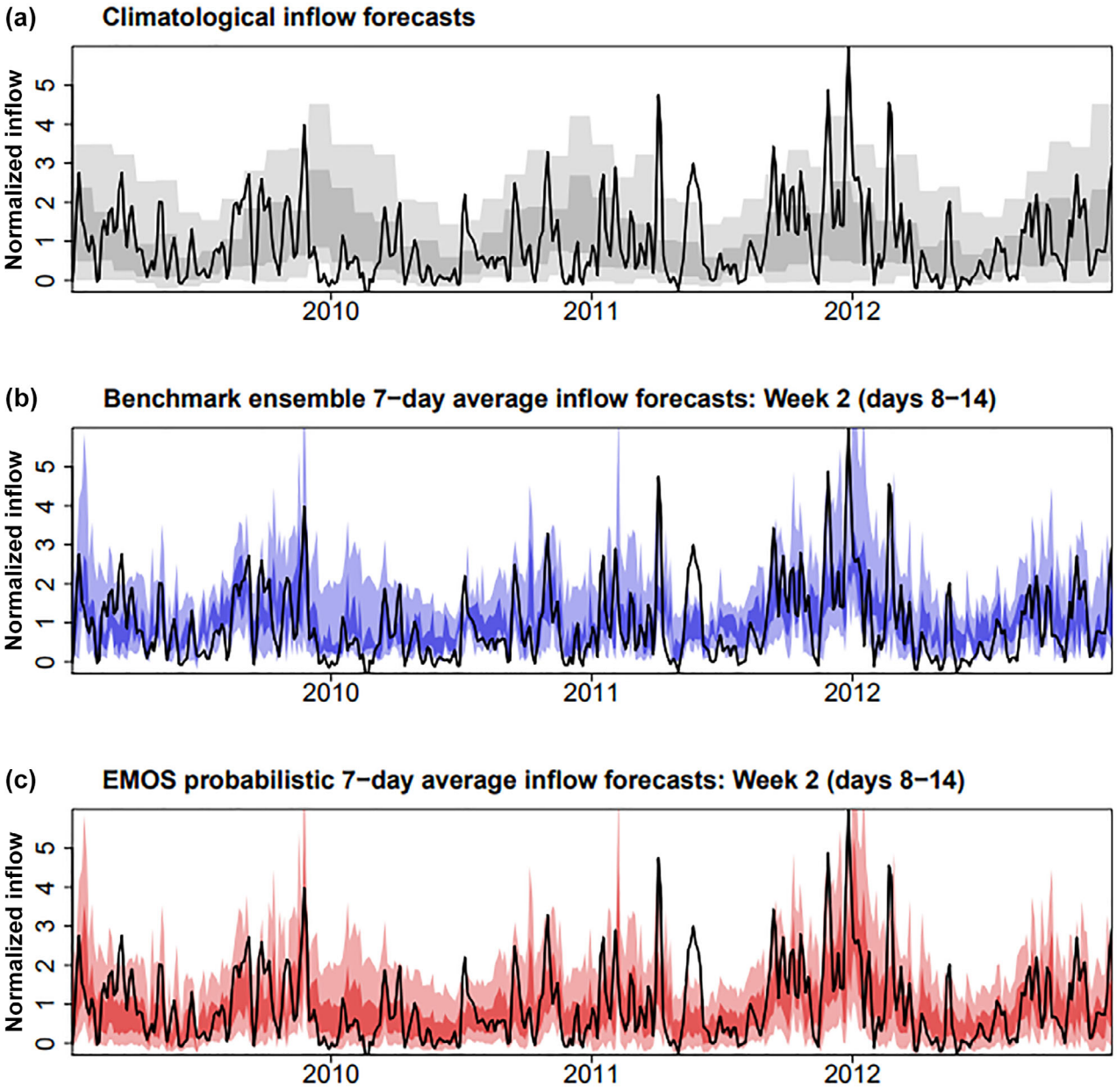


FIGURE 5 Time series of observed 7-day mean inflow rate for the case study reservoir. Inflow values are normalized by the annual mean inflow (a) climatological inflow forecasts (based on historical inflow observations). (b) Benchmark ensemble inflow forecasts for week 2 (days 8–14). (c) EMOS probabilistic forecasts for week 2. Shaded regions indicate the 50% and 90% prediction intervals

forecasts also demonstrate fair skill (fCRPSS = 0.03) for 6-week (days 1–42) average inflow forecasts, relative to the observed inflow climatology distributions (Figure 4d). It is clear that the skill of the S2S benchmark inflow forecasts is lower than the skill of the S2S precipitation predictions (Figure 4). This indicates that the linear regression model used to convert predicted precipitation into an inflow rate is imperfect.

Weeks 1 and 2 of the benchmark ensemble inflow forecasts largely capture the observed temporal variability of inflow at the case study reservoir (Figure 5). However, the ensemble inflow forecasts are under-dispersive. When comparing the observed inflow and the spread of ensemble members for the benchmark inflow forecasts, it is clear that the observed inflow frequently falls close to or outside the limits of the ensemble spread (Figure 5b). This is particularly common during periods when the observed inflow rate is high or low. When comparing the distribution of the observed inflow and forecast inflow, it is clear that the benchmark inflow forecasts under-predict the occurrence of both high and low inflow events (Figure 6).

The EMOS post-processing calibrates the inflow forecasts, and thus corrects the under-dispersion of the

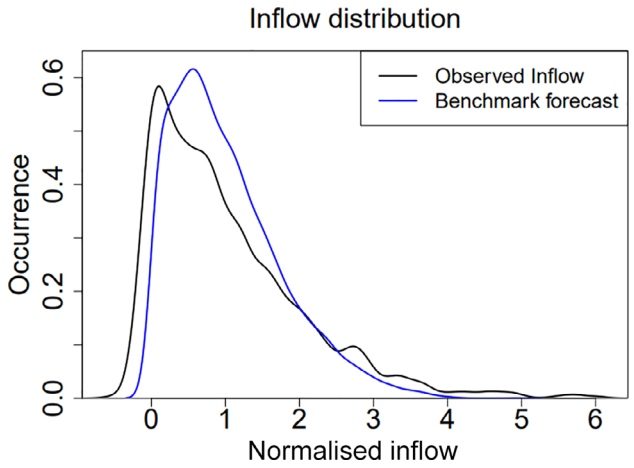


**FIGURE 6** Comparison of the observed distribution of weekly mean inflow rate (black) and week 1 (days 1–7) of the benchmark ensemble inflow forecasts (blue) over the 2009–2018 observation period. Inflow values are normalized by the annual mean inflow rate

benchmark S2S ensemble inflow forecasts. Hence, the observed inflow more frequently falls within the forecast distribution of the calibrated probabilistic inflow forecasts (Figure 5b,c). Post-processing also improves the probabilistic accuracy of the S2S inflow forecasts. Probabilistic accuracy means, for example, that the empirical proportion of observations falling below the 10% quantile of the forecast distribution should be 10%. Reliability diagrams show that the S2S probabilistic inflow forecasts perform well in this measure, with limited deterioration in probabilistic accuracy for increasing lead time (Figure 7). Importantly, post-processing does not correct errors that originate from the S2S ensemble weather predictions failing to predict an event, such as the high inflows during summer 2011 (Figure 8b,c).

As expected, due to improved calibration, post-processing greatly increases skill, relative to the benchmark inflow forecasts (Figure 4c,d). The fCRPSS for the S2S probabilistic inflow forecasts is significantly higher than the benchmark ensemble forecasts for all forecast horizons. The largest increase in skill score for the 7-day average inflow forecasts is during week 1 (days 1–7), where the fCRPSS increases by 0.14. For weeks 4–6 the fCRPSS increases by approximately 0.06. Importantly, 7-day average S2S probabilistic inflow forecasts hold fair skill relative to climatology for week 3 to week 6, with fCRPSS values between 0.01 and 0.04 (Figure 4a). Hence, post-processing breaks the forecasting skill horizon of 14 days lead time. Nonetheless, while significantly greater than zero, the skill scores for forecasts week 3 to week 6 are low and suggest a marginal improvement upon the use of an observed climatological inflow distribution.

Significantly greater skill is available for users interested in forecasts of the average inflow rate over periods longer than 7 days. Notably, the 6-week average probabilistic inflow forecasts (days 1–42) hold good skill relative to climatological forecasts (fCRPSS = 0.15), and are comparable in skill to the 7-day average probabilistic forecasts for week 2 (days 8–14) (Figure 4d and Figure 9). The skill of long-range average inflow forecasts decreases gradually with increasing forecast horizon, compared with a rapid reduction in forecast skill with increasing lead time for weekly average inflow forecasts (Figure 4). These forecasts are likely to be highly valuable to hydropower reservoirs with a large storage ratio (Turner et al., 2020).

## 5.4 | The use and value of S2S inflow forecasts for the hydropower sector

To deliver forecast value, the S2S inflow forecasts must be reliable and consistently improve water management decisions over an extended period of time. The forecasts should allow hydropower operators to optimize power generation to periods of peak energy prices, in order to achieve a higher average unit rate for the power generated.

In the stylized cost model framework, the value of the S2S probabilistic inflow forecasts follows similar patterns to those found for the forecast skill. The mean value of the 7-day average S2S probabilistic inflow forecasts decreases with increasing lead time (Figure 10a). The 7-day average probabilistic inflow forecasts for week 2 (days 8–14) deliver additional value of £2.20/MWh, relative to the use of climatological forecasts, when averaged over the full range of price differentials simulated. This value decreases to £1.40/MWh for week 6 (days 36–42) forecasts. When applying forecasts of longer duration, the 2-week mean (days 1–14) probabilistic inflow forecasts deliver an average additional value of £2.70/MWh, decreasing to £1.10/MWh for the 6-week (days 1–42) average forecasts (Figure 10b). Overall, we find that the S2S probabilistic inflow forecasts consistently deliver additional economic value, relative to the use of point climatological forecasts (Figure 10a,b). This conclusion holds for a range of forecast horizons up to 6 weeks ahead and a range of price differentials between peak and off-peak energy prices. This clearly demonstrates that the S2S probabilistic inflow forecasts are sufficiently reliable to improve water management decisions, consistent with the identification of statistical forecast skill (Figure 4).

The highest net unit rate achieved in the stylized cost model framework is through the application of S2S

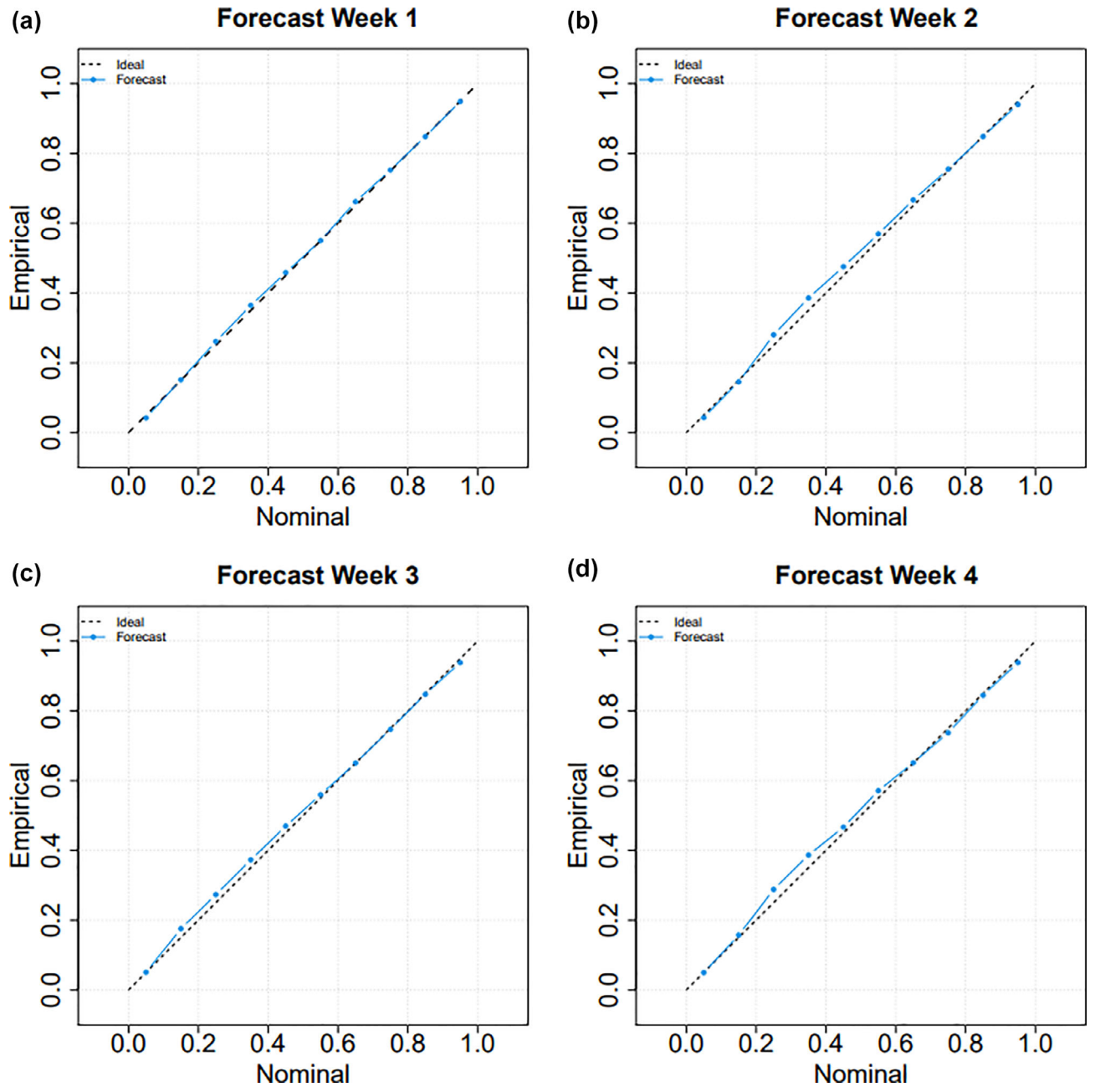


FIGURE 7 Reliability diagrams for weeks 1–4 of the EMOS probabilistic inflow forecasts. These show the empirical frequency at which observations (vertical axis) fall within the nominal forecast probability (horizontal axis). For perfect forecast reliability, the forecast probability should equal frequency of occurrence and the blue line should lie on the diagonal dashed line

probabilistic 6-week (days 1–42) average inflow forecasts (not shown), even though these forecasts deliver lower additional economic value compared with forecasts of shorter duration (Figure 10a,b). This is because we measure the additional economic value delivered by the probabilistic inflow forecasts relative to climatological forecasts. The average unit rate achieved by the climatological forecasts increases significantly for forecasts of longer duration. For example, a climatological forecast for the average inflow rate over a 6-week period will achieve a higher unit rate compared with using climatological forecasts of the average inflow rate over a single week. This is because there is significantly less interannual variability in the inflow rate averaged over an entire month, compared with a single week within that month. Thus, a climatological forecast for the entire month of January will be more accurate and result in smaller errors, compared with a climatological forecast for an individual week within January. The 6-week average inflow forecasts are most relevant to hydropower

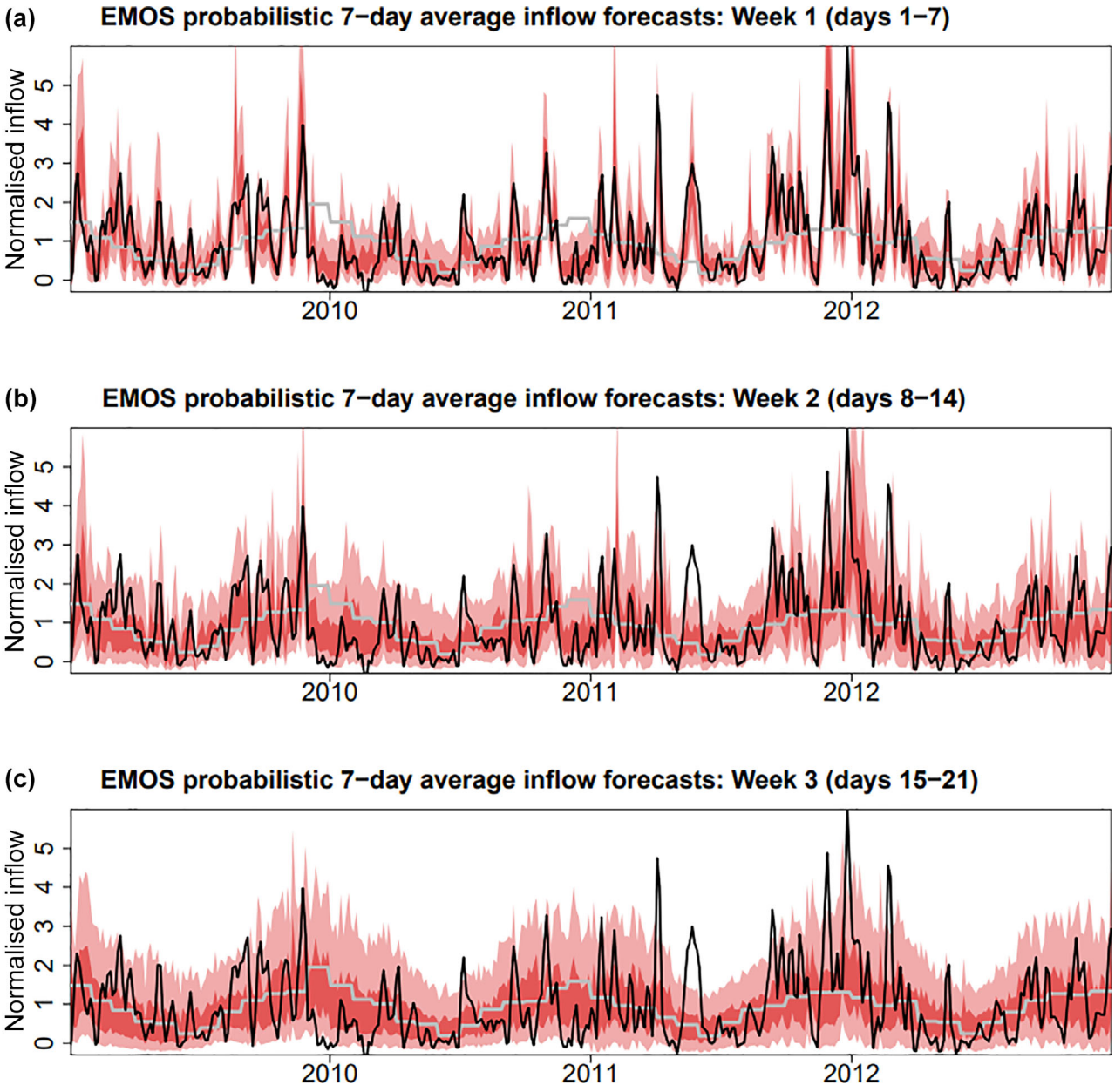


**FIGURE 8** Time series of observed weekly mean inflow rate (black) and EMOS probabilistic inflow forecasts for (a) week 1 (days 1–7), (b) week 2 (days 8–14), and (c) week 3 (days 15–21). Shaded areas indicate the 50% and 90% prediction intervals. Grey line shows the point climatological forecast, based on inflow observations. Inflow values are normalized by the annual mean inflow. Full 2009–2018 observation period is shown in Data S1

reservoirs with a large storage volume relative to annual mean inflow (Turner et al., 2020).

The stylized cost model framework reveals that the value of the S2S probabilistic inflow forecasts is non-linearly related to the price differential between peak and off-peak energy prices (Figure 10). Probabilistic forecasts deliver the largest increase in economic value, relative to point climatological forecasts, when the price differential is either very high (greater than £60/MWh) or low (less than £20/MWh). This result indicates that probabilistic inflow forecasts may become increasingly valuable if future energy prices are more volatile, as they are expected to in the coming years. When the difference between peak and off-peak energy prices is high, extending power generation into off-peak periods incurs a large opportunity cost. This means there is a high cost for taking precautionary action to keep the reservoir at its target water level, such as extending generation (i.e., discharge) into off-peak periods to accommodate forecast high inflows. In addition, there are substantial costs associated with inaccurate forecast information that result in deviations from the target water level. Under

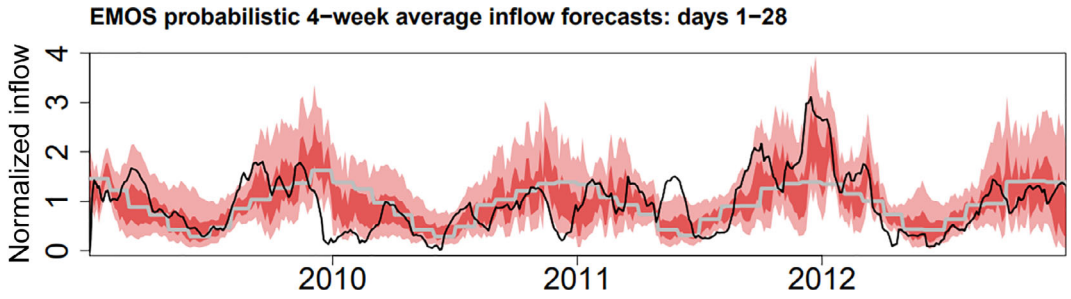


FIGURE 9 Time series of observed 28-day mean inflow at the case study reservoir (black line), and EMOS probabilistic 4-week (days 1–28) average inflow forecasts (red). Shaded areas indicate the 50% and 90% prediction intervals. Grey line shows the point climatological forecast based on inflow observations. Inflow values are normalized by the annual mean inflow. Full 2009–2018 observation period is shown in Data S1

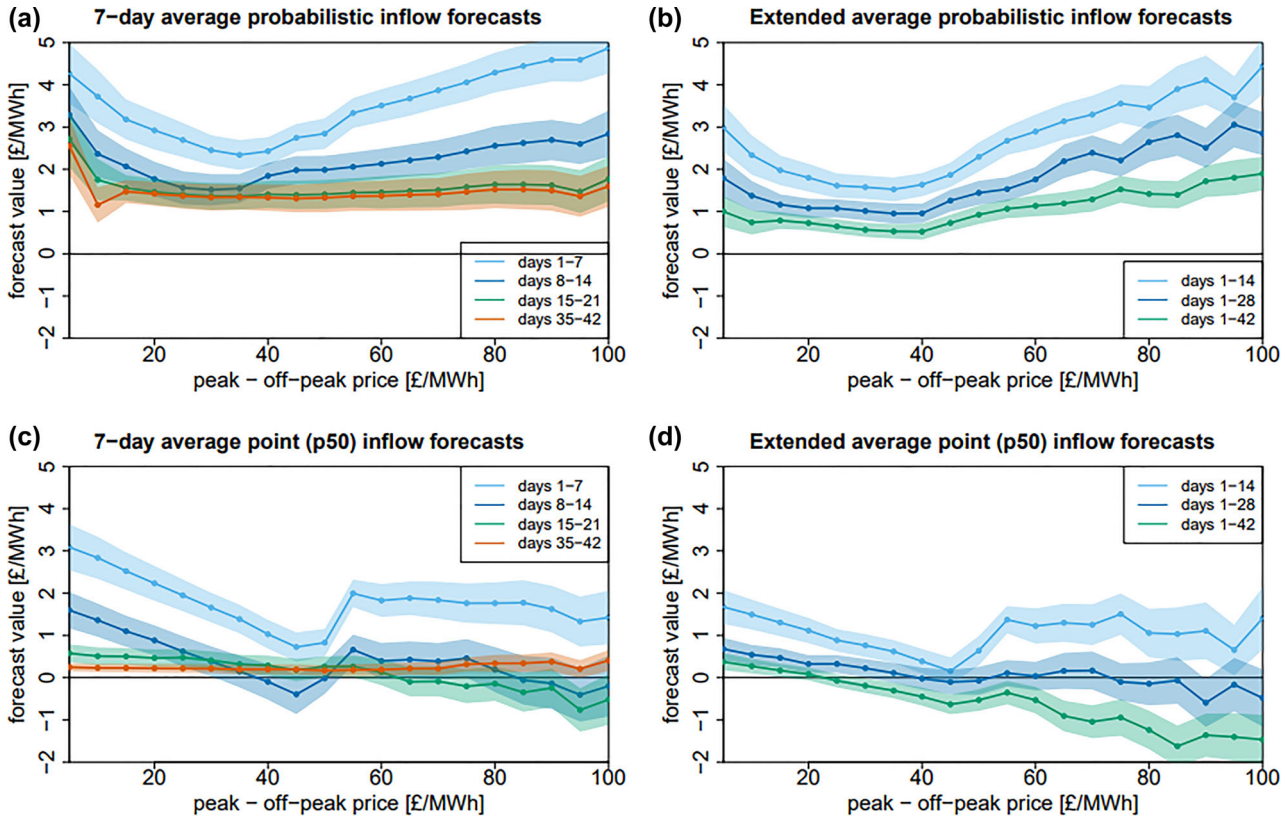


FIGURE 10 Economic value of (a)–(b) probabilistic and (c)–(d) point (p50/median) S2S inflow forecasts for different forecast horizons. Plots show the average additional economic value achieved, in terms of unit energy price, relative to point climatological forecasts. The average forecast value is measured over the entire historical observation record (2009–2018). Forecast value is simulated for a range of price differentials between peak and off-peak energy prices, while the peak price is kept constant at £50/MWh. Shaded areas indicate two standard errors (approx. 90% confidence interval), based on bootstrap resampling of data. See Data S1 for further details

these conditions, probabilistic (as opposed to point) forecasts are essential to guide optimal water management decisions. This is because probabilistic forecasts allow operators to fully consider risks posed by high-cost, low-probability outcomes.

The point (p50 median) inflow forecasts consistently deliver lower economic value compared with the probabilistic forecasts (Figure 10). Moreover, under certain conditions, the S2S point inflow forecasts deliver negative forecast value relative to the application of climatological forecasts (Figure 10c,d). Thus, the application of S2S point inflow forecasts can lead to poor water management decisions that result in economic losses, relative to a climatological forecast. This is true even for the 7-day average point inflow forecast for week 2 (days 8–14). In contrast with probabilistic inflow forecasts, point

forecasts hold lower economic value when the price differential is high. This is because point forecasts do not provide information on the possibility of high-cost, low-probability outcomes, which become increasingly important for higher price differentials. The point inflow forecasts deliver the greatest economic value when the price differential is low (less than £20/MWh). This is because power sold during off-peak periods will achieve a similar price to power sold during peak periods. For example, generation from a hydropower reservoir can be extended from 12 h per day to 24 h per day, to accommodate forecast high inflows, with limited financial implications. Hence, when the price differential is low, the cost of taking precautionary action is low. Similarly, there are limited costs associated with inaccurate forecast information that result in deviations from the target water level, unless a spill occurs.

It is important to note that the stylized cost model used here is a highly simplified representation of current operations. An important limitation is that we treat each forecast individually and do not account for possible accumulations of forecast errors over time. In addition, some of the economic value that is highlighted within the S2S probabilistic inflow forecasts is likely already realized in present operations through situational awareness of plant operators and meteorologists.

To summarize, in the stylized cost model framework, the S2S probabilistic inflow forecasts are found to consistently deliver increased economic value relative to point (p50 median) and climatological forecasts, in terms of the mean unit rate energy price achieved for the same water resource. This demonstrates that the S2S probabilistic inflow forecasts are sufficiently reliable to improve water management decisions. The value probabilistic forecasts are generally found to increase with higher price differentials between peak and off-peak energy prices, while the value of ensemble-median decreases under these situations and in some cases is negative. This demonstrates how the asymmetry of costs associated with forecast errors necessitates the use of probabilistic forecast information.

## 5.5 | Impact of atmospheric and seasonal variability on forecast skill

The seasonal variability of weather patterns and modes of atmospheric variability can affect forecast skill. Here, we evaluate the skill of the 7-day average S2S probabilistic inflow forecasts separately for the extended summer (April–September) and winter (October–March) seasons. We find that the inflow forecasts consistently hold higher skill during winter than summer, across all lead times (Figure 11a). However, the difference in skill is not always significant. The largest difference in skill scores between the summer and winter seasons is during forecast week 2 (days 8–14). Here, wintertime forecasts hold good skill, while summer forecasts demonstrate fair skill (Figure 11a). Skill scores for weeks 3–6 of the summer forecasts are not significantly greater than zero. Similar

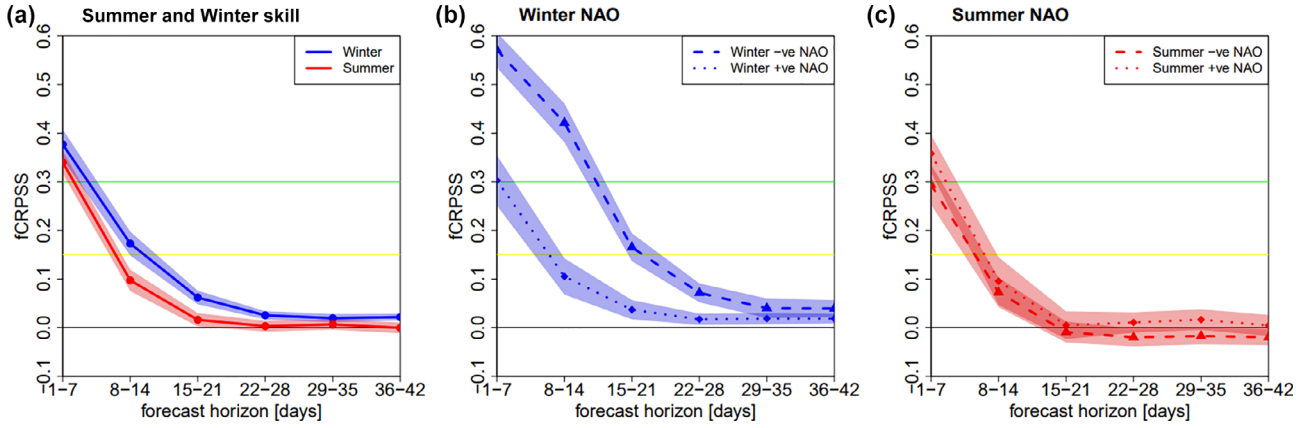


FIGURE 11 Impact of seasonal and atmospheric variability on the skill of S2S probabilistic inflow forecasts for different forecast horizons (Table 1). (a) Forecast skill for summer (April–September, red) and winter (October –march, blue). (b) Impact of North Atlantic oscillation on winter inflow forecast skill. (c) Impact of North Atlantic oscillation on summer inflow forecast skill. Positive NAO periods correspond to an NAO index >+0.4 (dotted lines), and a negative NAO corresponds to an NAO index <−0.4 (dashed lines). Forecast skill is measured using the fair continuous ranked probability skill score. Forecast skill is assessed over the 2009–2018 observation period. Shaded areas indicate +/− 2 standard deviations. Forecast skill is measured using the fair continuous ranked probability skill score, using the SpecsVerification package in R. an fCRPSS value >0 is classed as fair forecast skill, progressing to good forecast skill for values above 0.15 (yellow line) and very good for values above 0.3 (green line)

patterns are found for the benchmark inflow forecasts and S2S ensemble weather predictions (not shown).

Several recent studies have also found higher skill for streamflow and precipitation forecasts in the United Kingdom and Europe during winter months, compared with summer (Arnal et al., 2018; Bell et al., 2017; Svensson, 2015; Weisheimer & Palmer, 2014). During the winter season, teleconnections have been identified between weather patterns in the tropics, such as the Madden–Julian oscillation (MJO), that occur via linkages with the polar stratospheric vortex and provide sources of predictability in the eastern North Atlantic region in the S2S range (Cassou, 2008; Lee et al., 2019; Vitart, 2017, 2017). This could explain the higher skill of the S2S inflow forecast during winter.

We further evaluate the skill of the S2S probabilistic inflow forecasts for different phases of the NAO. The highest skill scores for the 7-day average S2S probabilistic inflow forecasts are during winter months with a negative NAO index (Figure 11). The largest difference in skill score for the different phases of the NAO during winter is for the forecast horizon of week 2 (days 8–14), when forecasts during negative phases of the NAO hold very good skill and forecasts during positive phases hold fair skill (Figure 10b). Negative phases of the NAO index during winter are often associated with colder and drier than average conditions over large parts of the United Kingdom. Thus, these periods typically coincide with low inflow and streamflow rates. The winters of 2009–2010 and 2010–2011 are two examples where negative phases of the NAO (Scaife, 2014) coincided with prolonged, exceptionally low inflow rates at the case study reservoir. The 4-week average (days 1–28) S2S probabilistic inflow forecasts perform well during the winters of 2009–2010 and 2010–2011, predicting clear signals of below-average inflows (Figure 9). The availability of reliable S2S inflow forecasts would likely have greatly assisted the operations of the hydropower sector during these periods and improved water management.

Reasons for high S2S skill scores during wintertime NAO− may be due to the stable and persistent nature of these weather patterns, with low precipitation over western Scotland (Lavers et al., 2013). In addition, teleconnections between an active MJO over the western Pacific and wintertime NAO− are also enhanced during La Niña years, such as 2010–2011, providing additional sources of predictability on S2S timescales (Lee et al., 2019). However, the forecast skill scores are measured relative to climatology, and the climatological forecasts perform particularly poorly during these low winter inflow events. This is compounded by our short observational time series. Thus, higher skill scores may be due to a poorer climatological forecast, rather than better S2S inflow forecast. We note that the primary source of skill for the 28-day average inflow forecast is from the first 2 weeks of the S2S forecasts (Figure 4). There is limited signal of below-average inflows from the 7-day average inflow forecasts during week 3 (days 15–21) (Figure 8c).

These analyses of how the phase of the NAO impacts the skill of inflow forecasts must be treated with caution, due to the relatively short observational time series. Furthermore, there is a limited number of positive and negative NAO events to evaluate in each season. This is reflected by the large standard deviation for the fCRPSS values (Figure 11).

The skill of inflow forecasts for the case study reservoir is substantially lower during summer months compared with winter. During summer, no significant skill is available beyond week 2 (days 8–14) (Figure 11). Weisheimer and Palmer (2014) describe the ECMWF seasonal precipitation forecasts as being "dangerously useless" during dry summers in the United Kingdom, because they may provide forecast users with inaccurate forecast information. There is a greater fraction of convective precipitation in the United Kingdom during summer months compared with winter (Lavers et al., 2013). Due to the smaller-scale processes involved, convective precipitation is more challenging to model and forecast compared with large-scale precipitation. The random initiation of convective storms means these features cannot forecast reliably beyond a few days ahead. Even for short-range forecasts, it is difficult to resolve the exact path of these storms and local topographic interactions, due to the relatively coarse horizontal and vertical resolution of the numerical models used (Anghileri et al., 2019; Scheuerer, 2014). The S2S probabilistic forecasts clearly struggle to predict high summer inflow events, such as summer 2011 (Figure 8). This may also help to explain the lower skill of the S2S probabilistic inflow forecasts during summer.

The fact the S2S probabilistic inflow forecasts hold higher skill in winter months, compared with summer, is important for hydropower applications. Water levels in reservoirs are typically highest during winter, because of higher seasonal inflow rates. This means that the greatest risk of spillages is during winter, and thus reliable forecast information is essential. As reservoir water levels are lower during summer months, there are lower risks of spills. This means that costs associated with incorrect forecast information are lower during summer.

## 6 | CONCLUSIONS

In this study, we present a new method of producing S2S probabilistic inflow forecasts for hydropower reservoirs

using S2S ensemble weather predictions, without the application of a hydrological model. We generate probabilistic inflow forecasts for a case study reservoir in Scotland, and evaluate predictions for 11 different forecast horizons, covering periods up to 6 weeks ahead. The forecasts are evaluated using observed reservoir inflow rates.

We generate S2S inflow forecasts by training a linear regression model for the observed inflow rate on to historical S2S predictions of precipitation. This regression model is applied to convert ensemble S2S predictions of precipitation rates into an ensemble of S2S inflow rate forecasts. We subsequently apply EMOS post-processing techniques that build upon the method of Scheuerer (2014), to produce calibrated probabilistic forecasts of inflow.

The S2S probabilistic forecasts for the 7-day mean inflow rate hold fair skill up to 6 weeks ahead (days 35–42), relative to climatological forecasts. While significantly greater than zero, the skill for forecasts week 3 to week 6 is relatively low and suggests marginal improvement upon the skill of observed climatological inflow distributions. However, greater skill is realized by considering forecasts of the average inflow rate over periods longer than 7-days. For example, probabilistic forecasts of the average inflow rate over the next 6 weeks (days 1–42) demonstrate good skill relative to climatological forecasts, and hold comparable skill to 7-day average probabilistic inflow forecasts for 2 weeks ahead (days 8–14). These long-range average inflow rate forecasts are likely to be valuable for hydropower reservoirs with a large storage volume relative to annual mean inflow rate.

We develop a stylized cost model to assess the potential economic value of the S2S probabilistic inflow forecasts may deliver to the hydropower sector. This demonstrates that the probabilistic forecasts are sufficiently reliable to consistently improve water management decisions, relative to climatological forecasts, delivering a higher average unit rate for the power generated. In the model framework, the value of the probabilistic forecasts increases for higher price differentials between peak and off-peak energy prices. Reducing the probabilistic forecasts to a single point value, based on the median of the probabilistic forecast, significantly reduces the forecast value and in some cases results in negative forecast value. This demonstrates how probabilistic forecasts are essential to allow operators to fully consider the risks of high-cost, low-probability outcomes. We conclude that, if energy prices become more volatile, S2S probabilistic inflow forecasts are likely to become increasingly valuable to the hydropower sector.

While the focus of this study is on the hydropower sector in Scotland, the methods outlined here will be relevant for other water management applications, such as forecasting streamflow in rivers, public water availability, and early warning systems for floods and droughts.

## AUTHOR CONTRIBUTIONS

**Robert M. Graham:** Conceptualization (supporting); data curation (lead); formal analysis (lead); investigation (equal); methodology (equal); project administration (supporting); resources (equal); software (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Jethro Browell:** Conceptualization (lead); data curation (supporting); formal analysis (supporting); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (equal); supervision (lead); validation (supporting); writing – review and editing (equal). **Douglas Bertram:** Conceptualization (supporting); funding acquisition (supporting); investigation (supporting); methodology (supporting); supervision (supporting); writing – review and editing (supporting). **Christopher J. White:** Conceptualization (supporting); funding acquisition (supporting); investigation (supporting); methodology (supporting); supervision (supporting); writing – review and editing (supporting).

## DATA AVAILABILITY STATEMENT

The 2019 hindcast dataset for the ECMWF extended-range forecasts was accessed from the S2S prediction project (http://s2sprediction.net/) database hosted by the ECMWF (https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/). ERA5 reanalysis files were accessed from the NCAR-UCAR Research Data Archive (https://rda.ucar.edu/). Daily NAO index was accessed from the National Weather Service Climate Prediction Center (https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml). Forecast skill scores were calculated using the package 'SpecsVerification' in R. The normalized inflow record for the case study reservoir will be published together with this manuscript.

## ORCID

*Robert M. Graham* https://orcid.org/0000-0003-0008-1886
*Jethro Browell* https://orcid.org/0000-0002-5960-666X

## REFERENCES

Ahmad, S.K. & Hossain, F. (2019) A generic data-driven technique for forecasting of reservoir inflow: application for hydropower

maximization. *Environmental Modelling and Software*, 119, 147–165. https://doi.org/10.1016/j.envsoft.2019.06.008

Anghileri, D., Monhart, S., Zhou, C., Bogner, K., Castelletti, A., Burlando, P. et al. (2019) The value of subseasonal Hydrometeorological forecasts to hydropower operations: how much does preprocessing matter? *Water Resources Research*, 55, 10159–10178. https://doi.org/10.1029/2019WR025280

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

Baker, S.A., Wood, A.W. & Rajagopalan, B. (2020) Application of post-processing to watershed-scale sub-seasonal climate forecasts over the contiguous U.S. *Journal of Hydrometeorology*, 21 (5), 971–987. https://doi.org/10.1175/jhm-d-19-0155.1

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Båserud, L., Izsák, B. et al. (2022) Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *International Journal of Climatology*, 42(2), 727–747. https://doi.org/10.1002/joc.7269

Barnston, A.G. & Livezey, R.E. (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115, 1083–1126. https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2

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, 4681–4691. https://doi.org/10.5194/hess-21-4681-2017

Bloomfield, H.C., Brayshaw, D.J. & Charlton-Perez, A.J. (2020) Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications*. 27(1). https://doi.org/10.1002/met.1858

Bloomfield, H., Brayshaw, D., Gonzalez, P. & Charlton-Perez, A. (2021) Sub-seasonal forecasts of demand, wind power and solar power generation for 28 European countries. *Earth System Science Data Discussions*, 13(5), 2259–2274. https://doi.org/10.5194/essd-2020-312

Branković, Č., Palmer, T.N., Molteni, F., Tibaldi, S. & Cubasch, U. (1990) Extended-range predictions with ECMWF models: time-lagged ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, 116, 867–912. https://doi.org/10.1002/qj.49711649405

Brayshaw, D.J., Halford, A., Smith, S. & Jensen, K. (2020) Quantifying the potential for improved management of weather risk using sub-seasonal forecasting: the case of UKtelecommunications infrastructure. *Meteorological Applications*, 27. https://doi.org/10.1002/met.1849

Buizza, R. & Leutbecher, M. (2015) The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, 141, 3366–3382. https://doi.org/10.1002/qj.2619

Cassou, C. (2008) Intraseasonal interaction between the madden–Julian oscillation and the North Atlantic oscillation. *Nature*, 455, 523–527. https://doi.org/10.1038/nature07286

Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X. et al. (2021) Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. *Atmospheric Research*, 250, 105365. https://doi.org/10.1016/j.atmosres.2020.105365

European Centre for Medium-Range Weather Forecasts. (2017). updated monthly. ERA5 Reanalysis. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6X34W69. Accessed 01 Sept 2020.

Ferro, C.A.T. (2014) Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140, 1917–1923. https://doi.org/10.1002/qj.2270

Gneiting, T. & Katzfuss, M. (2014) Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. & Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562. https://doi.org/10.1038/NCLIMATE3338

Hall, R.J. & Hanna, E. (2018) North Atlantic circulation indices: links with summer and winter UKtemperature and precipitation and implications for seasonal forecasting. *International Journal of Climatology*, 38, e660–e677. https://doi.org/10.1002/joc.5398

Hall, R.J., Hanna, E., Scaife, A.A., Hanna, E., Jones, J.M. & Erdélyi, R. (2017) Simple statistical probabilistic forecasts of the winter NAO. *Weather and Forecasting*, 32, 1585–1601. https://doi.org/10.1175/WAF-D-16-0124.1

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, 2023–2039. https://doi.org/10.5194/hess-22-2023-2018

Hersbach, H, Bell, B, Berrisford, P, Hirahara, S, Horányi, A, Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Ho, L.T.T., Dubus, L., De Felice, M. & Troccoli, A. (2020) Reconstruction of multidecadal country-aggregated hydro power generation in Europe based on a random Forest model. *Energies*, 13, 1786. https://doi.org/10.3390/en13071786

Hu, X. & Yuan, W. (2021) Evaluation of ERA5 precipitation over the eastern periphery of the Tibetan plateau from the perspective of regional rainfall events. *International Journal of Climatology*, 41, 2625–2637. https://doi.org/10.1002/joc.6980

Huertas-Hernando, D. & Co-authors. (2017) Hydro power flexibility for power systems with variable renewable energy sources: an IEA task 25 collaboration. *Wiley Interdisciplinary Reviews: Energy and Environment*, 6, e220. https://doi.org/10.1002/wene.220

Johnson, S.J. & Co-authors. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12, 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Khouja, M. (1999) The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, 27, 537–553. https://doi.org/10.1016/S0305-0483(99)00017-1

Klemm, T. & McPherson, R.A. (2017) The development of seasonal climate forecasting for agricultural producers. *Agricultural and Forest Meteorology*, 232, 384–399. https://doi.org/10.1016/j.agrformet.2016.09.005

Lavers, D., Prudhomme, C. & Hannah, D.M. (2013) European precipitation connections with large-scale mean sea-level pressure

(MSLP) fields. *Hydrological Sciences Journal*, 58, 310–327. https://doi.org/10.1080/02626667.2012.754545

Lee, R.W., Woolnough, S.J., Charlton-Perez, A.J. & Vitart, F. (2019) ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophysical Research Letters*, 46, 13535–13545. https://doi.org/10.1029/2019GL084683

Lledó, L., Torralba, V., Soret, A., Ramon, J. & Doblas-Reyes, F.J. (2019) Seasonal forecasts of wind power generation. *Renewable Energy*, 143, 91–100. https://doi.org/10.1016/j.renene.2019.04.135

Magnusson, J., Nævdal, G., Matt, F., Burkhart, J.F. & Winstral, A. (2020) Improving hydropower inflow forecasts by assimilating snow data. *Hydrology Research*, 51, 226–237. https://doi.org/10.2166/nh.2020.025

Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N. & Doblas-Reyes, F.J. (2020) Choices in the verification of S2S forecasts and their implications for climate services. *Monthly Weather Review*, 148, 3995–4008. https://doi.org/10.1175/MWR-D-20-0067.1

Merryfield, W.J. & Co-authors. (2020) Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, 101, E869–E896. https://doi.org/10.1175/BAMS-D-19-0037.1

Messner, J.W., Pinson, P., Browell, J., Bjerregård, M.B. & Schicker, I. (2020) Evaluation of wind power forecasts—an up-to-date view. *Wind Energy*, 23, 1461–1481. https://doi.org/10.1002/we.2497

Ødegård, H.L., Eidsvik, J. & Fleten, S.-E. (2019) Value of information analysis of snow measurements for the scheduling of hydropower production. *Energy Systems*, 10, 1–19. https://doi.org/10.1007/s12667-017-0267-3

Olaniyan, E., Adefisan, E.A., Oni, F., Afiesimama, E., Balogun, A.A. & Lawal, K.A. (2018) Evaluation of the ECMWF subseasonal to seasonal precipitation forecasts during the peak of West Africa monsoon in Nigeria. *Frontiers in Environmental Science*, 6, 1–15. https://doi.org/10.3389/fenvs.2018.00004

Orsolini, Y.J., Kindem, I.T. & Kvamstø, N.G. (2011) On the potential impact of the stratosphere upon seasonal dynamical hindcasts of the North Atlantic oscillation: a pilot study. *Climate Dynamics*, 36, 579–588. https://doi.org/10.1007/s00382-009-0705-6

Rigby, R.A. & Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society*, 54, 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Scaife, A.A. & Co-authors. (2014) Skillful long-range prediction of European and north American winters. *Geophysical Research Letters*, 41, 2514–2519. https://doi.org/10.1002/2014GL059637

Scheuerer, M. (2014) Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society*, 140, 1086–1096. https://doi.org/10.1002/qj.2183

Sene, K. (2016) *Hydrometeorology*. Switzerland: Springer International Publishing.

Soret, A. & Co-authors. (2019) Sub-seasonal to seasonal climate predictions for wind energy forecasting. *Journal of Physics Conference Series*, 1222, 012009. https://doi.org/10.1088/1742-6596/1222/1/012009

Svensson, C., Brookshaw, A., Scaife, A.A., Bell, V.A., Mackay, J.D., Jackson, C.R. et al. (2015) Long-range forecasts of UK winter hydrology. *Environmental Research Letters*, 10(6), 064006. https://doi.org/10.1088/1748-9326/10/6/064006

Svensson, C. (2016) Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues. *Hydrological Sciences Journal*, 61, 19–35. https://doi.org/10.1080/02626667.2014.992788

Tarek, M., Brissette, F.P. & Arsenault, R. (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24, 2527–2544. https://doi.org/10.5194/hess-24-2527-2020

Turner, S.W.D., 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. https://doi.org/10.5194/hess-24-1275-2020

van der Wiel, K., Bloomfield, H.C., Lee, R.W., Stoop, L.P., Blackport, R., Screen, J.A. et al. (2019) The influence of weather regimes on European renewable energy production and demand. *Environmental Research Letters*, 14(9), 094010. https://doi.org/10.1088/1748-9326/ab38d3

Vannitsem, S., Wilks, D. & Messner, J.W. (2018) *Statistical Postprocessing of ensemble forecasts*. Amsterdam, Netherlands: Elsevier.

Vautard, R. (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Monthly Weather Review*, 118, 2056–2081. https://doi.org/10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2

Vitart, F. (2004) Monthly forecasting at ECMWF. *Monthly Weather Review*, 132, 2761–2779. https://doi.org/10.1175/MWR2826.1

Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. https://doi.org/10.1002/qj.2256

Vitart, F. (2017) Madden—Julian oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, 143, 2210–2220. https://doi.org/10.1002/qj.3079

Vitart, F. & Co-authors. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98, 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Vitart, F. & Robertson, A.W. (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1, 3. https://doi.org/10.1038/s41612-018-0013-0

Wang, L. & Robertson, A.W. (2019) Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems. *Climate Dynamics*, 52, 5861–5875. https://doi.org/10.1007/s00382-018-4484-9

Wanner, H., Brönnimann, S., Casty, C., Gyalistras, D., Luterbacher, J., Schmutz, C. et al. (2001) North Atlantic oscillation – concepts and studies. *Surveys in Geophysics*, 22, 321–381. https://doi.org/10.1023/A:1014217317898

Weisheimer, A. & Palmer, T.N. (2014) On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, 11, 20131162. https://doi.org/10.1098/rsif.2013.1162

White, C.J. & Co-authors. (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, 24, 315–325. https://doi.org/10.1002/met.1654

Wood, A.W. & Lettenmaier, D.P. (2008) An ensemble approach for attribution of hydrologic prediction uncertainty. *Geophysical Research Letters*, 35, L14401. https://doi.org/10.1029/2008GL034648

Woolnough, S.J. (2019) *The madden-Julian oscillation*. Sub-Seasonal to Seasonal Prediction: Elsevier, pp. 93–117.

Zhang, F., Sun, Y.Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H. et al. (2019) What is the predictability limit of Mid-latitude weather? *Journal of the Atmospheric Sciences*, 76, 1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Graham, R. M., Browell, J., Bertram, D., & White, C. J. (2022). The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting. *Meteorological Applications*, *29*(1), e2047. https://doi.org/10.1002/met.2047