Reservoir Status & Outlooks

Methods & Application

Nathan Rickards, Helen Baron & Rishma Chengot

Issue number 1.0
March 2024

UK Centre for Ecology & Hydrology

NATIONAL CAPABILITY FOR GLOBAL CHALLENGES
International science for net zero plus

Contents

1. Introduction

The hydrological cycle is a dynamic phenomenon occurring through a range of Earth processes associated with atmosphere, land, oceans, and life on Earth. A key element of this cycle is water storage in the form of lakes and reservoirs. Both lakes and reservoirs form essential components of the hydrological and biogeochemical water cycles as they are able to store, retain, clean, and provide water consistently (Crétaux *et al.*, 2016). This is especially pertinent in light of increasing demands on the water sector due to a growing population and a changing future climate (Donchyts *et al.*, 2022).

Despite this, the impact of water body storages on the system as a whole is not fully understood, largely due to a lack of information on the amount of water being stored and how this changes over time (Busker *et al.*, 2019). This is especially the case for water stored in reservoirs. In many countries the privatization of water companies means that operating rule curves and associated data are rarely available to the public, and there remains a lack of both spatially and temporally consistent data for these water bodies (Salwey *et al.*, 2023; Steyaert and Condon, 2023).

Reservoir numbers have increased significantly over the last half century (Biswas *et al.*, 2021); there currently exists more than 16 million reservoirs worldwide with a combined storage capacity of over 8 million cubic metres (Lehner et al. 2011), and over half of the world's large river systems are currently impacted by dams (Nilsson *et al.*, 2005). Reservoirs are largely built for the supply and management of water resources; they play an integral role in hydroelectricity, irrigation, and domestic water supply needs, and are key for the management of water resources (Gao, Birkett and Lettenmaier, 2012). Available water resources for human use and water stress, both in the present and future periods, can only be meaningfully evaluated when human alterations to the hydrological cycle are taken into account (Biemans *et al.*, 2011). It is therefore vital to monitor reservoirs to support the evaluation of water body storage and the management of water resources (Gusyev, 2015). Tools and associated metrics for this type of evaluation allow practitioners to better understand both anthropogenic impacts and climate change influences on behaviours of water supply and demand (Haddeland *et al.*, 2014).

1.1 Aims of the report

The establishment of a standardised approach to represent reservoir status and outlooks would be a useful tool in supporting water practitioners, as well as providing a clear and meaningful pathway for the dissemination of water resources information to the general public. As such, initiatives such as the UK Hydrological Summary, the UK Hydrological Outlook, and the WMO HydroSOS, would benefit from the development and application of such tools to facilitate their current outputs around water resources status, both present and future.

To address these needs, this report will:

- build upon previous work by Rickards et al. (2022) to explore the potential of reservoir status metrics for the reporting of water storage status and outlooks;
- explore the use of forecasting methods to predict future changes in water body storage.

This report contains a review of the current state of the science with respect to reservoir metrics and forecasting methods in Section 2. Different storage metrics are then explored in Section 3, with discussion of their relative strengths and potential visual presentations, to improve communication of reservoir storage status. In Section 4 we investigate data-driven models for prediction and forecasting of reservoir storage at a range of lead times for selected reservoirs over the UK.

2. State of the science

2.1 Reservoir metrics

There is wide recognition that a standardised index can help support the evaluation of water body storage and the monitoring and management of water resources (Gusyev, 2015). Such metrics are already in widespread use for hydrological variables, including for precipitation via the Standardised Precipitation Index (SPI), for groundwater fluctuations via the Groundwater Resilience Index (GRI) (Chinnasamy, Maheshwari and Prathapar, 2018), and for streamflow via the Standardised Streamflow Index (SSI) (Barker *et al.*, 2015). Metrics are often more appropriate figures to report than absolute values in watershed management, largely due to the fact that certain data is often of a sensitive nature and can be highly emotive (Boongaling, Faustino-Eslava and Lansigan, 2018; Avisse *et al.*, 2017; Vieira Valadão *et al.*, 2021). This is especially pertinent for data associated with reservoir resources due to how water distribution is prioritised and subsequently allocated.

An example of a metric already in use for reservoir storage is the Reservoir Storage Index (RSI) as used by Tiwari et al. (2019). The RSI, based on the SPI metric developed by McKee et al. (1993), was calculated to study the temporal variability in reservoir storage in India. The metric requires an input of monthly reservoir storage, and was used to forecast reservoir storage anomalies in a study to assess the severity and duration of water resource deficits in the dry season. Similarly, Gusyev et al. (2016) devised an approach based on the SPI methodology to analyse the status of reservoir storage in drought conditions: the Standardised Reservoir Supply Index (SRSI). The metric was implemented as part of a drought assessment framework of standardised indices in the Pampanga (Philippines), Solo (Indonesia), and Chao Phraya (Thailand) river basins, specifically for the assessment of socio-economic drought. Differing slightly to the approach of Tiwari et al. (2019), the metric requires inputs of reservoir inflows alongside available water storage, and compares SRSI with a Standardised Discharge Index (SDI). Both approaches standardise anomalies against a reference period, over either a monthly or seasonal accumulation period.

Another approach to quantifying anomalous values in water body storage is the percentage difference in storage away from a monthly long-term mean (referred to here as Difference from the Monthly Mean (DMM)), as utilised in the UKCEH Hydrological Summary (NRFA, 2022; UK Met Office, 2022). This method again allows for a trend in water storage volumes to be identified and given as a metric, in this case the percentage difference from a monthly average. This approach can be used with actual values or an alternative representative of water storage e.g. percentage of total capacity, as is the case in the Hydrological Summary. From these the anomaly values per month or season for a reservoir can be reported. The user can therefore calculate the metric in different units, arguably making it more versatile and potential more meaningful for practitioners, than the RSI approach.

Building on the above approaches, UKCEH have produced a metric based on the methodology set out for streamflow status for the World Meteorological Organisation's (WMO) HydroSOS initiative (World Meteorological Organization, 2020). Here, percentiles are calculated using the Weibull distribution to rank the water storage of the current month against an historical baseline period. From this, a percentile range can be categorised based on its rate of occurrence, as derived from the historical period. By doing this, the current month's storage is considered alongside all other historical values for that particular month as opposed to being referenced against a long-term average, as in the RSI and DMM methodologies. The Environment Agency (EA) use a similar methodology in their assessment of reservoir status in England as part of their national monthly reports, where percentile bands are categorised based on the occurrence of a particular storage range seen over the duration of the historic (or baseline) period (Environment Agency, 2024).

A suitable reference period must be established for all the above methodologies, often requiring data which covers a relatively long time period. For the SPI metric, The World Meteorological Organisation (WMO) recommends at least 20-30 years of monthly values, with 50-60 being preferred (World Meteorological Organisation, 2012). There is currently no recognised standard length of reference period for assessing trends in water storage status. The current UKCEH Hydrological Summary uses 25 years, based purely on data availability across all of its monitored sites, the same justification as Gusyev (2015), who use between 20-25 years of data. Whilst a longer time series better captures long-term trends and gives greater confidence in any metric produced, limitations arise as a result of both the length and quality of data available. Users therefore must make a subjective decision as to a suitable length and completeness of reservoir storage data in order to produce a meaningful and relevant storage metric.

2.2 Forecasting

Given the importance of reservoirs to water resources, it is important to have a reliable forecast for reservoir status to ensure efficient operation of individual reservoirs and the wider water resource system (Peñuela, Hutton and Pianosi, 2020; Ahmad and Hossain, 2019), particularly during critical periods such as extreme drought (Turner *et al.*, 2017) and flooding events (Wang *et al.*, 2012; Zarei *et al.*, 2021). This is a current and vital issue for many regions worldwide, and will become increasingly important under changing climate and growing demand for water and hydropower (Gleick, 2003; Gleick *et al.*, 2013). As such, there has been a lot of research on forecasting reservoir status (i.e. reservoirs storage, inflow, outflow, level, or storage anomaly) at different lead-times and with different methods.

Change in reservoir storage is the difference between the sum of incoming flows (inflow, rainfall) and the sum of outgoing flows (reservoir release or diversions, evaporation, recharge to groundwater), and is dominated by inflow and reservoir release/diversions (hereafter referred to as outflow). While inflow can be predicted using typical hydro-meteorological variables and methods, outflow can be difficult to predict if there are irregular diversions, for example, the irregular industrial usage of water from reservoirs in the Brazos rivers basin in Texas (Fernando, Zhu and Negusse, 2017).

Global hydrological and land-surface models that include reservoir representation can be used in forecast mode to predict reservoir storage, e.g. H08 (Hanasaki *et al.*, 2008), WaterGap (Alcamo *et al.*, 2003; Döll, Kaspar and Lehner, 2003), PCR-GLOBWB

(Wada, van Beek and Bierkens, 2012), VIC (Droppers *et al.*, 2020), LPJmL (Biemans *et al.*, 2011), CWATM (Burek *et al.*, 2019), and JULES (Clark *et al.*, 2011; Best *et al.*, 2011). Benefits of this method include global coverage, the ability to forecast at virtually any lead-time, and no requirements for historical data on reservoir storage. However, these models rely on generic release schemes to simulate reservoir operations, which inevitably results in inaccuracies when locally important operational drivers are neglected. More flexible reservoir simulations are being developed to improve global reservoir simulations, including empirically derived storage-release functions incorporated into large-scale hydrological models (Yassin *et al.*, 2019; Turner, Doering and Voisin, 2020), and fuzzy reservoir operational rules determined by artificial neural networks (Coerver, Rutten and van de Giesen, 2018).

This approach is, however, limited by the flow biases present in the models (Turner, Doering and Voisin, 2020). To avoid accuracy issues arising from generic schemes and flow biases, catchment scale models can be used to simulate reservoir storage (or level) as these can incorporate local reservoir operation rules and be calibrated for a given catchment (Gronewold *et al.*, 2011; Hughes, Birkinshaw and Parkin, 2021; Kim *et al.*, 2020; Zhao *et al.*, 2016). The gains in accuracy by using a small-scale model are balanced by the reduction in spatial coverage and the increased data requirements (e.g. reservoir operational rules, streamflow for calibration, etc.). Hydrological models are sometimes used to forecast reservoir inflow rather than reservoir storage (Anghileri *et al.*, 2016; Xu *et al.*, 2021) which is then used to optimise reservoir operations (Ficchì *et al.*, 2016).

As an alternative to hydrological models, data-driven models have become popular for forecasting. These include empirically defined storage-release functions (Yassin *et al.*, 2019; Turner, Doering and Voisin, 2020), autoregressive integrated moving average (ARIMA) models (Valipour, Banihabib and Behbahani, 2013; Ibañez *et al.*, 2021), various Artificial Neural Network (ANN) models (Valipour, Banihabib and Behbahani, 2013; Zhang *et al.*, 2018; Ehsani *et al.*, 2016; Ibañez *et al.*, 2021; Zhu *et al.*, 2020; Coerver, Rutten and van de Giesen, 2018) and Recurrent Neural Networks (RNN), including the Long Short-Term Memory (LSTM) model (García-Feal *et al.*, 2022; Özdoğan-Sarıkoç *et al.*, 2023), random forest (RF) and related algorithms (Wang and Wang, 2020; Sapitang *et al.*, 2020; Hong *et al.*, 2020), as well as other machine learning (ML) models (Raghavendra. N and Deka, 2014; Zarei *et al.*, 2021; Yang *et al.*, 2017). While some studies forecast reservoir storage directly (Kim *et al.*, 2022; Özdoğan-Sarıkoç *et al.*, 2023), others forecast reservoir levels (Wang and Wang, 2020; Ibañez *et al.*, 2021; Sapitang *et al.*, 2020; Zhu *et al.*, 2020; Castillo-Botón *et al.*, 2020), which can be used to estimate reservoir storage or as a proxy for storage. Data-driven models have also been used to forecast reservoir inflow (Valipour, Banihabib and Behbahani, 2013; Yang *et al.*, 2017; Hong *et al.*, 2020; Choong and El-Shafie, 2015; Zarei *et al.*, 2021), reservoir outflow (Yang *et al.*, 2016; García-Feal *et al.*, 2022), and reservoir storage anomalies (Tiwari and Mishra, 2019).

The main advantage of these models is their ability to capture reservoir operations without prior knowledge of operational rules. However, they do require sufficiently long observational records of the target variable (e.g. reservoir storage, level, inflow, or anomaly), and the performance of these models can be reduced by anthropogenic effects (Özdoğan-Sarıkoç *et al.*, 2023). Other explanatory variables employed in these models include current and lagged hydro-meteorological variables such as rainfall, air temperature, relative humidity, short and long-wave radiation, upstream and downstream flow, snow depth, snowmelt, and climate phenomenon indices such as the Oceanic Niño Index. A few studies have included demand-related data, for example, irrigation releases (Ibañez *et al.*, 2021), legally mandated water allocations (Yang *et al.*, 2016), and water used to generate electricity (Castillo-Botón *et al.*, 2020). When using data-driven models it is important to remember that they may not perform well in conditions beyond the range seen in the training data, and that they are vulnerable to future changes (e.g. variation in demand, increasing climatic extremes, new upstream reservoirs, etc.). This can be mitigated to some extent by training a "universal" model which uses timeseries data from multiple reservoirs along with static characteristics, thus exposing the model to a greater range of conditions in the training data. This idea has been demonstrated by Kratzert et al. (2019) for rainfall runoff modelling. Another potential limitation of ML methods is the lack of interpretability inherent to "black-box" models (some ML models, such as neural networks, fall into this category).

A combination of physical-based and data-driven models have been explored for use in hydrological forecasting, particularly in streamflow (Ng *et al.*, 2023; Hunt *et al.*, 2022; Degenne *et al.*, 2024; Kraft *et al.*, 2022). These hybrid models show promise for modelling complex systems with increased accuracy and interpretability and could be considered for modelling reservoirs, particularly if multiple variables are being predicted (e.g. reservoir inflow, storage, and outflow).
An important consideration when forecasting reservoir status is the forecast lead-time. The majority of the forecasts performed with ML models have one-step ahead forecasts, with a lead-time of a day or a week depending on the data, but for operational purposes long-term or multi-step forecasting is often more useful. Long-term forecasts of reservoir level were considered in Castillo-Botón et al. (2020) using a persistence-based approach, and multiple ML models were used in Ibañez et al. (2021) to forecast reservoir level at lead-times of 1, 30, 90, and 180 days.

3. Storage status and metrics

A methodology that allows for the reporting of reservoir storage status without the need to divulge storage values negates many of the sensitivities surrounding water supplies and current water stocks. Here we look at two anomaly-based methodologies, as well as a percentile-based categorisation method. The Wimbleball reservoir is used as a case study for all metric calculations. This reservoir, located in Devon, UK, and primarily used for water supply, has a period of record spanning more than 30 years and very few missing data observations (Figure 1).

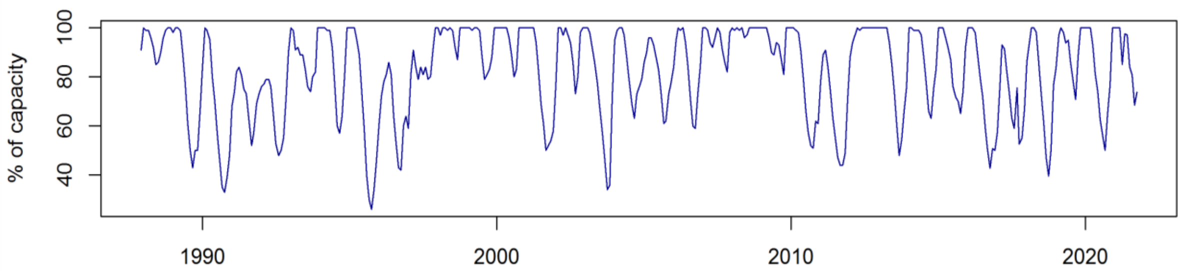

Figure 1: Percentage of reservoir capacity as reported for Wimbleball reservoir, UK; 1989-2020.

3.1 Data and methods

To derive water storage status, a time series of reservoir volumes (or levels and corresponding waterbody dimensions) is generally required: for our case study reservoirs we use observed data sourced from the representative water authority via the NRFA (2024). To extend this work to other reservoirs, we would ideally source observed data from, for example, a national hydrological monitoring program or relevant water authority. However, this may not always be possible and so alternative data sources may need to be considered. These alternative data may also be required for the process of infilling, where data gaps or uncertainties exist in observed time series. Potential alternative data sources are reviewed in Rickards et al. (2022), along with a more extensive catalogue of freely available reservoir data collated through the Copernicus In Situ (COINS) project (Rickards *et al.*, 2023).

Difference from the long-term monthly mean (DMM)

The percentage difference in storage away from the long-term monthly mean (DMM) is the metric currently utilised by UKCEH to report the status of reservoir storage in the Hydrological Summary (NRFA, 2024). This method calculates the average percentage storage for each calendar month in a reference period, and then presents current storage as a percentage difference away from the corresponding average monthly means.

The metric is calculated for the Wimbleball reservoir, UK, in Figure 2. The example displays the metric for October through to December 2020, using a baseline period of 1989-2019.

Table 1 Anomaly, current and long-term monthly average values (percent storage) for the Wimbleball reservoir, October-December 2020.

Location	Reservoir	Oct 2020	Nov 2020	Dec 2020
Exmoor, UK	Wimbleball	Anom.= -0.6 Cur.= 64.9 Av. = 65.5	Anom.= 3.2 Cur.= 76.3 Av. = 73	Anom.= 17.5 Cur.= 100 Av. = 82.4

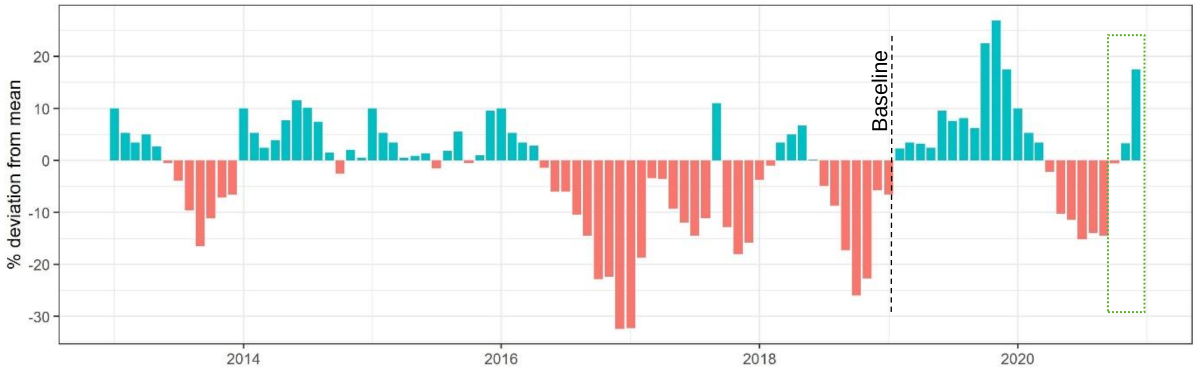

Figure 2: Anomaly values calculated using the DMM at Wimbleball reservoir; dotted line shows the end of the baseline period (1989-2019); green box shows the period which metrics are being calculated for in Table 1.

As displayed in Table 1 and Figure 2, storage at Wimbleball in October 2020 was slightly below the long-term average for the month. However, both November and December show slight increases on the long-term means for the respective months, at 3.2% and 17.5%, respectively. Although Table 1 gives both the current and long-term average storage in percent alongside the anomaly metric, it may be preferable for users to report the anomaly metric only for reservoirs where there is more sensitivity around data sharing.

The calculation of the DMM metric is simple and the outputs are relatively straighforward to interpret. The UKCEH Hydrological summary currently reports this status metric in both tabular and graphical formats, as seen in Figure 2 and Table 1.

Reservoir Storage Index (RSI)

The Reservoir Storage Index (RSI) (McKee, Doesken and Kleist, 1993; Tiwari and Mishra, 2019) uses long-term monthly storage data to calculate a distribution function of reservoir storage, which is then transformed to a normal distribution so that an RSI of zero describes the normal storage for the given location, month and accumulation period (Svoboda, 2012). Deviations from this (both positive and negative) are then expressed in terms of standard deviations. More extreme values indicate that these deviations are more severe, but also less likely to occur.

Standardised indices can be calculated for month and season e.g. for 1 and 3 month accumulation periods. A one-month accumulation period was used here, so that the results could be compared to the current period as reported by the UKCEH Hydrological Summary.

There is a further metric, the Standardised Reservoir Storage Index (SRSI) (Gusyev, 2015), which is similar to the RSI but requires the additional input of reservoir inflow. This metric has therefore not been utilised here, due to the extra data requirement which is not available for many reservoirs.

The RSI is based on the SPI metric as defined by McKee et al. (1993). In the SPI, McKee et al. (1993) define metric categories to describe drought severity. These categories can be amended to describe reservoir storage, for example: mild reduction in stock (-1 to 0), moderate reduction in stock (-1.5 to -1), severe reduction in stock (-2 to -1.5) and extreme reduction in stock (less than -2). As the original metric is concerned with droughts, anything above 0 is considered to be wetter than normal; in the case of RSI, this indicates higher stock levels than normal.

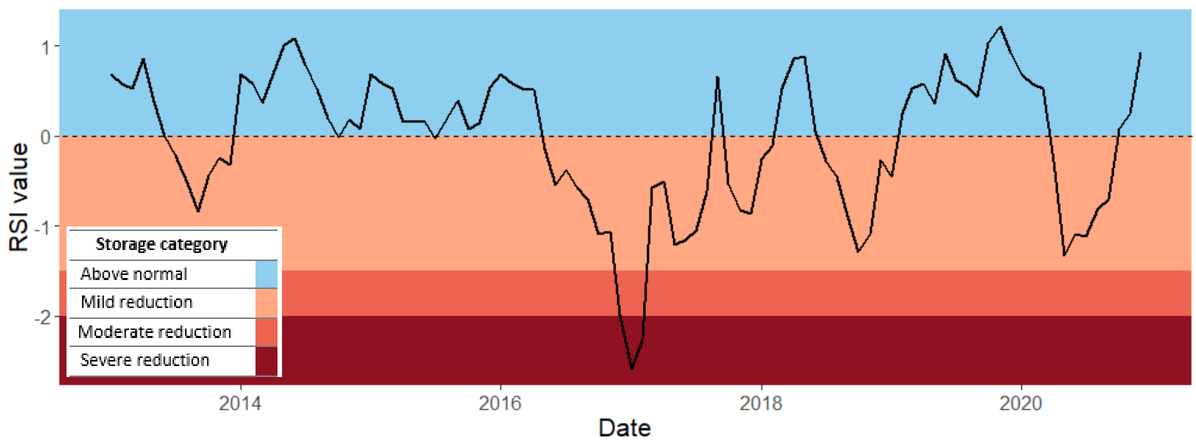

Figure 3: RSI values for the Wimbleball reservoir.

The RSI was calculated for the Wimbleball reservoir and presented in Figure 3. The general trends of the DMM are replicated by the RSI metric (Figure 2 and Figure 3), for example, at the end of 2016 the DMM shows >30% difference from the long-term monthly mean and the RSI is <-2, indicating an extreme, but rare, level of reduction in reservoir volume. The unit used by this method (standard deviation away from the mean) makes it less intuitive from a water management perspective, although this could be desirable where data is highly sensitive.

Rank percentiles approach

Another approach to categorising reservoir storage follows that used by the WMO for the reporting of streamflow status in HydroSOS,and is based upon ranking observed historical storage into percentiles. This has recently been applied by UKCEH for the WMO HydroSOS metric for streamflow status and is currently being explored as a possible methodology for use in the UK Hydrological Summary for reservoir storage. In this method, reservoir storage values from a defined baseline period are sorted and ranked per month. Where the lowest values are equal (e.g. dead storage) then the maximum rank is used, and where the highest values are equal (e.g. maximum reservoir storage) then the lowest rank is used. Where other values match i.e. anywhere between minimum and maximum values, the average rank is used. This is to ensure that multiple recordings of minimum and/or maximum capacity are assigned the same rank value. From here, the Weibull distribution is applied to the ranked values, which is used to assign a percentile to each record of storage. User defined percentile ranges can then be used to indicate where the current storage sits in terms of historic values, as seen in Table 2. The categories applied here are based upon the methodology in Barker et al. (2022).

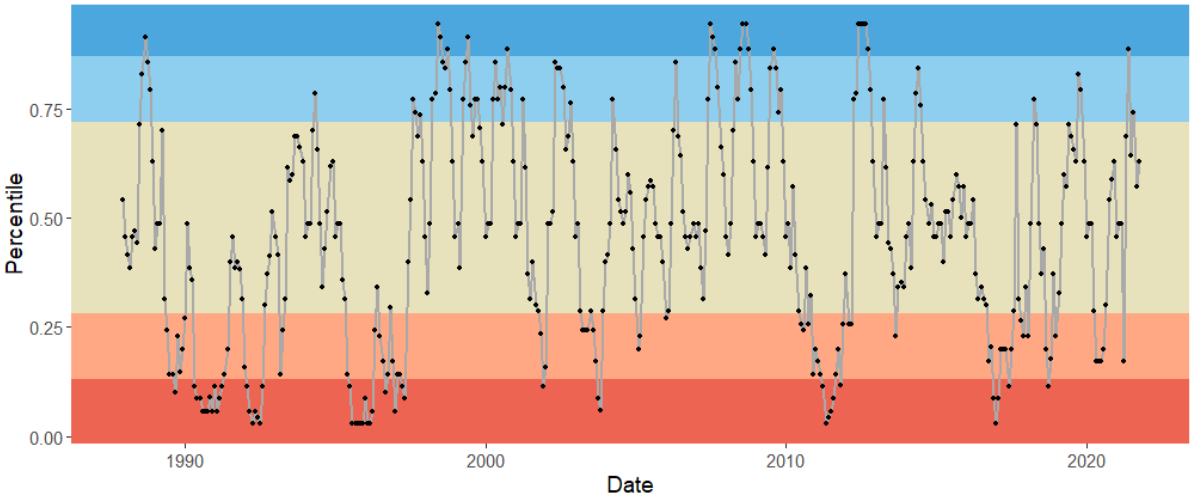

Figure 5: Timeseries of rank percentiles for Wimbleball, 1989-2021.

Figure 4 displays the time series of rank percentile values for the Wimbleball. This clearly demonstrates where values are deemed to be outside of a normal range for a particular month. In Figure 5 a selection of reservoir storages in the UK are displayed for a given month. Where the current storage sits compared to the baseline period is indicated via the colour of the circles, with the percentage storage of each reservoir given within its corresponding circle.

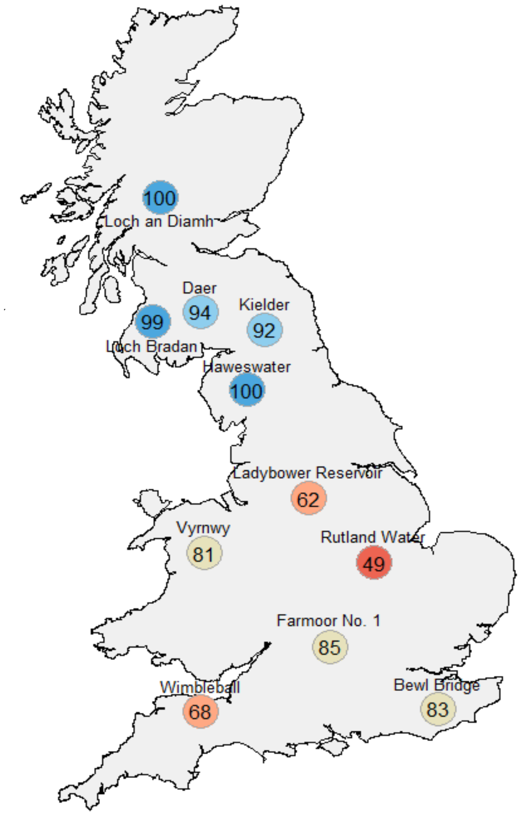

Table 2: Categories assigned to percentiles.

Storage category	Percentile range (%)
Notably high	87 – 100
Above normal	72 - 87
Normal range	**28 - 72**
Below normal	13 - 28
Notably low	0 - 13

Figure 6: Spatial output for the rank percentile method for selected UK reservoirs.

Such a methodology avoids the requirement of a long-term monthly mean, and instead ranks the value for a current month against all previous reported storage values. This approach is more robust in handling extreme historical values, in that these will be ranked as 'outliers' as opposed to skewing the long-term mean in either direction, especially where data records are short.

3.2 Summary: status metrics

The DMM, RSI and rank percentile metrics are all useful indicators of reservoir status, given the provision of appropriate data. Whilst some thought needs to be given to the amount of good quality data available, their use in facilitating monthly storage updates without the need to provide actual volumes could help facilitate the dissemination and communication of data around reservoir storage and status.

All metrics have the potential to be displayed in graphical, tabular or spatial (map) formats, and the choice of metric will therefore be determined by the needs of the stakeholder and availability of data. Whilst all metrics will tell a similar story, the unit given should be focussed around meeting the needs the target audience. For example, the RSI uses standard deviations away from a long-term average, which is ideal for showing general trends and indicating the rarity of the more extreme storage surpluses and deficits. However, it may be less useful for water users and practitioners wanting to know what this means in more relatable terms. In this situation, the DMM and rank percentile approach are more appropriate.

Regardless of the metric being used, consideration should be given to how these data are presented. For example, spatial outputs such as Figure 5 are extremely useful for capturing the current status of a number of reservoirs, and an approach which places a current status into a category of historical occurrence is a useful visual to help place current conditions into context. This, alongside a timeseries approach for individual reservoirs, is both informative and practical for a variety of users, without the need to divulge absolute values which may be regarded as sensitive.

4. Exploring Forecasting

In previous work several data-driven models were explored to scope their utility for forecasting reservoir status (Rickards and Baron, 2022). The aim of that study was to explore the skill of simple, data-driven methods for forecasting reservoir status at one to six month lead times. The SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors) and RF (Random Forest) models were selected since they are partially interpretable, relatively simple, and established in hydrological forecasting. Results from Rickards et al. (2022) showed that, while the SARIMAX model was relatively accurate, particularly at shorter lead times, the RF model likely requires more fine-tuning to produce a reliable forecast.

This work builds on the previous study, with a more in-depth exploration of RF models for forecasting storage in several case-study reservoirs in the UK. The RF type models were chosen for this work for several reasons: they are relatively simple models which can be built and applied with an intermediate knowledge of machine learning methods; they are partially interpretable, providing feature importance scores which can be used as a sense-check of the model and to provide insight into the physical processes driving reservoir storage change; they are well established in the field of hydrological prediction (Wang and Wang, 2020; Sapitang *et al.*, 2020; Hong *et al.*, 2020) and have been shown to have some skill in forecasting reservoir storage with potential for greater accuracy (Rickards and Baron, 2022).

4.1 Random Forests

A RF is made up of a collection of decision trees. A decision tree can predict the value of a target variable by learning simple decision rules inferred from the data features (explanatory variables) (Li *et al.*, 1984), see Figure 5 for an example decision tree. A RF model is an ensemble of decision trees (Figure 6), each of which learns from a random subset of features, with the final prediction being the average of each trees' prediction (Criminisi, 2011). A RF model produces a more accurate forecast compared to an individual decision tree but is more complex and less interpretable than a single tree model.

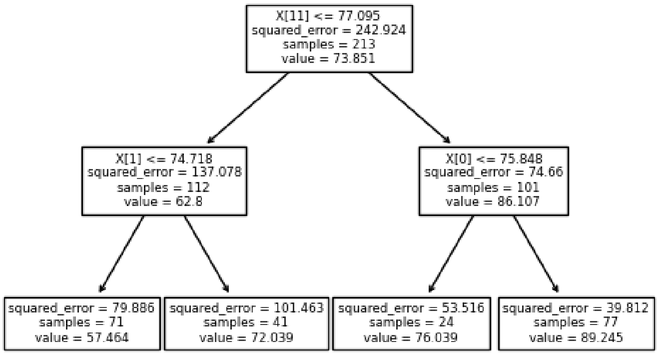

Figure 6: Example decision tree, with features X[i] for i=0 to 11, and continuous target value.

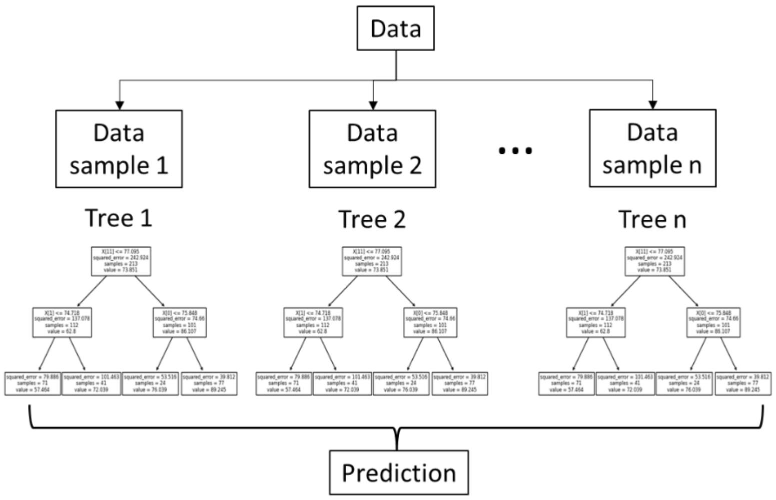

Figure 7: Visualisation of a Random Forest model.

RF type models are not specifically time series models, so lagged variables are generally used as input features. Multi-step forecasts can be made using either a direct method (i.e. a different model is developed for each lead-time), or a recursive approach (i.e. using a single model, with prior forecasts used as input for the next lead-time).

Method

The following steps were taken to produce a RF type model for forecasting, using the Python package Scikit-learn v1.3.0 (Pedregosa *et al.*, 2011):

- Reservoir selection: six reservoirs were chosen from the 39 time-series datasets of reservoir storage that were available. These reservoirs cover a range of geographies and characteristics, as detailed in Table 3.
- Data selection: monthly storage was available for each reservoir from 1988 to present day[1]. This was supplemented with monthly precipitation and temperature from the HadUK gridded climate data (Met Office *et al.*, 2018), chosen to match the climate data used in the UK hydrological forecasting (UK Centre for Ecology & Hydrology, 2023). These data series were then split into a training and testing set for the model to learn and predict on respectively, with the initial 75% of data points in the training set and the final 25% in the testing set.
- Model selection: a range of RF and related model types were tested on subsets of the training data, with the best performing model for each reservoir chosen. The full set of trialled models are detailed in Table A1 in the Appendix.
- Feature engineering and selection: various lagged and averaged variables were produced from the monthly storage, precipitation and temperature data series, the most influential of these (as determined by the feature importance property of the model) were then taken forward for each reservoir model. For more detail on this step see Table A2 and Table A3 in the Appendix**.**
- Hyperparameter tuning: the hyperparameter space for the selected model was explored using a random search method, with training and testing scores

[1] Reservoir storage were sourced from the representative water authority via the NRFA (2024).

inspected to check for overfitting. Overfitting is where the model fits the training data too well, so performs well on the training data but not the test data, and can be a risk when the training data is small (as it is in this study).

- Prediction and model evaluation: for each reservoir, the selected model with the chosen hyperparameters predicted reservoir storage for the testing period at one, three, and six month lead times, using a recursive approach for multi-step prediction. These predictions were evaluated against the observed storage, and compared to a baseline ARIMA model with the structure ARIMA$(1,0,1)(2,1,0)_{12}$.
- Forecast: using the historical analogue method employed in the UK hydrological forecasting, the models produced an ensemble of forecasts for reservoir storage over one and three month periods.

Table 3: Summary of reservoirs considered.

Reservoir Name	Latitude	Longitude	Catchment area (km^2)	Capacity (hm^3)	Mean depth (m)	Reservoir type	Year opened	Description
Ardingly	51.04	-0.09	22.77	5	6.1	Impounding	1977	In West Sussex, operated by South East Water; provides water and regulates water flows in the River Ouse.
Wimbleball	51.07	-3.47	28.79	21	13.1	Impounding	1978	On Exmoor; shared between South West Water (river regulation for abstraction) and Wessex Water (direct supply). Impounds the Haddoe; tributary of the Exe.
Clywedog	52.47	-3.60	48.92	45	20.1	Impounding	1956	Tallest concrete dam in UK. Outflows; for regulation of the Severn; via Clywedog.
Grafham	52.29	-0.28	15.92	55	9.7	Non-impounding	1964	Operated by Anglian Water; abstracts from River Great Ouse.
Rutland	52.66	-0.61	72.76	117	10.9	Non-impounding	1975	Largest surface area of any UK reservoir. Abstraction from Welland at Tinwell/Stamford and from Nene at Wansford.
Kielder Water	55.18	-2.46	242.42	199	18.3	Impounding	1982	Largest capacity reservoir in UK. England's largest HEP plant; generating power from the compensation flow to the North Tyne. Underpins Kielder Transfer Scheme which can transfer water to Wear & Tees catchments in times of drought.

The predictions for each model were compared to the observed data, and model performance was assessed using the coefficient of determination (R^2), defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where y_i are observed values, $\hat{y}_i$ are predicted values, $\bar{y}$ is the mean of the observed values. For R^2, the closer the value is to one, the better the model is performing.

Results and discussion

The prediction skill of the RF models at 1, 3 and 6 month lead time for the various reservoirs is presented in Table 3, along with the baseline ARIMA model. Predicted reservoir storage at a 1 month lead time is compared to observed storage for the test period in Figure 7.

The RF models perform better for almost all reservoirs at all lead times compared to the baseline ARIMA model (see Table 4), although this gap is reduced somewhat if the ARIMA predictions are capped at 100% reservoir storage (as the ARIMA model has a tendency to predict above 100% during high storage periods). The ARIMA models perform best on reservoirs with a clear seasonal storage pattern, such as Ardingly and Wimbleball, but perform poorly on reservoirs where this pattern is less pronounced, such as Clywedog. The RF models performs well for most of the reservoirs, even when the seasonal pattern is not strong, but demonstrate lower predictive skill for Kielder Water compared to the other reservoirs, possibly due to the erratic storage pattern exhibited in that reservoir. Kielder Water is the largest of the selected reservoirs, and has a high level of artificial management (see Table 4).

There are several periods of low reservoir storage that are evident across all the reservoirs (except Kielder Water): the autumns of 2018 and 2022, a result of particularly dry summers in 2018 and 2022 (see Figure 7). While the dry period of 2018 is well captured by both RF and ARIMA models, the ARIMA models in particular overpredict storage for most reservoirs in the 2022 dry period, and the RF models overpredict for Ardingly, Wimbleball, and (to some extent) Grafham. During localized dry periods such as autumn 2020 in Ardingly, RF models tend to overpredict, whereas ARIMA performs better. This may be due to the localised nature of the models employed here, i.e. the RF models only receive information on climatic conditions over their catchment area, but in reality reservoir operations will depend on regional and national conditions. It may also be due to the limited number of low storage events available for the model to train on, for example, storage at Wimbleball drops to 17.8% in the October of 2022, while the minimum storage in the training data for Wimbleball is 26% (October 1995). Both of these limitations may be addressed by combining all the reservoir data, along with static catchment characteristics, into a universal model. This approach has shown promise for data-driven rainfall-runoff modelling (Kratzert *et al.*, 2019).

All models perform better at shorter lead times, with a greater decrease in skill between 1 and 3 months, compared to 3 and 6 months for most reservoirs. The RF models perform well for most of the example reservoirs at 1 month lead time (except perhaps Kielder Water), and about half of the reservoirs at a longer lead time.

Table 4: Performance of the RF and ARIMA models for reservoir storage prediction at 1, 3 and 6 month lead time.

Reservoir	1 month		3 months		6 months	
	RF R^2	*ARIMA R^2*	*RF R^2*	*ARIMA R^2*	*RF R^2*	*ARIMA R^2*
Ardingly	0.875	0.801	0.782	0.400	0.734	0.307
Wimbleball	0.908	0.796	0.852	0.655	0.808	0.615
Clywedog	0.843	0.476	0.708	0.338	0.688	0.334
Grafham	0.802	0.527	0.535	0.095	0.355	-0.191
Rutland	0.778	0.718	0.420	0.438	0.180	0.477
Kielder Water	0.529	0.363	0.311	0.064	0.257	0.009

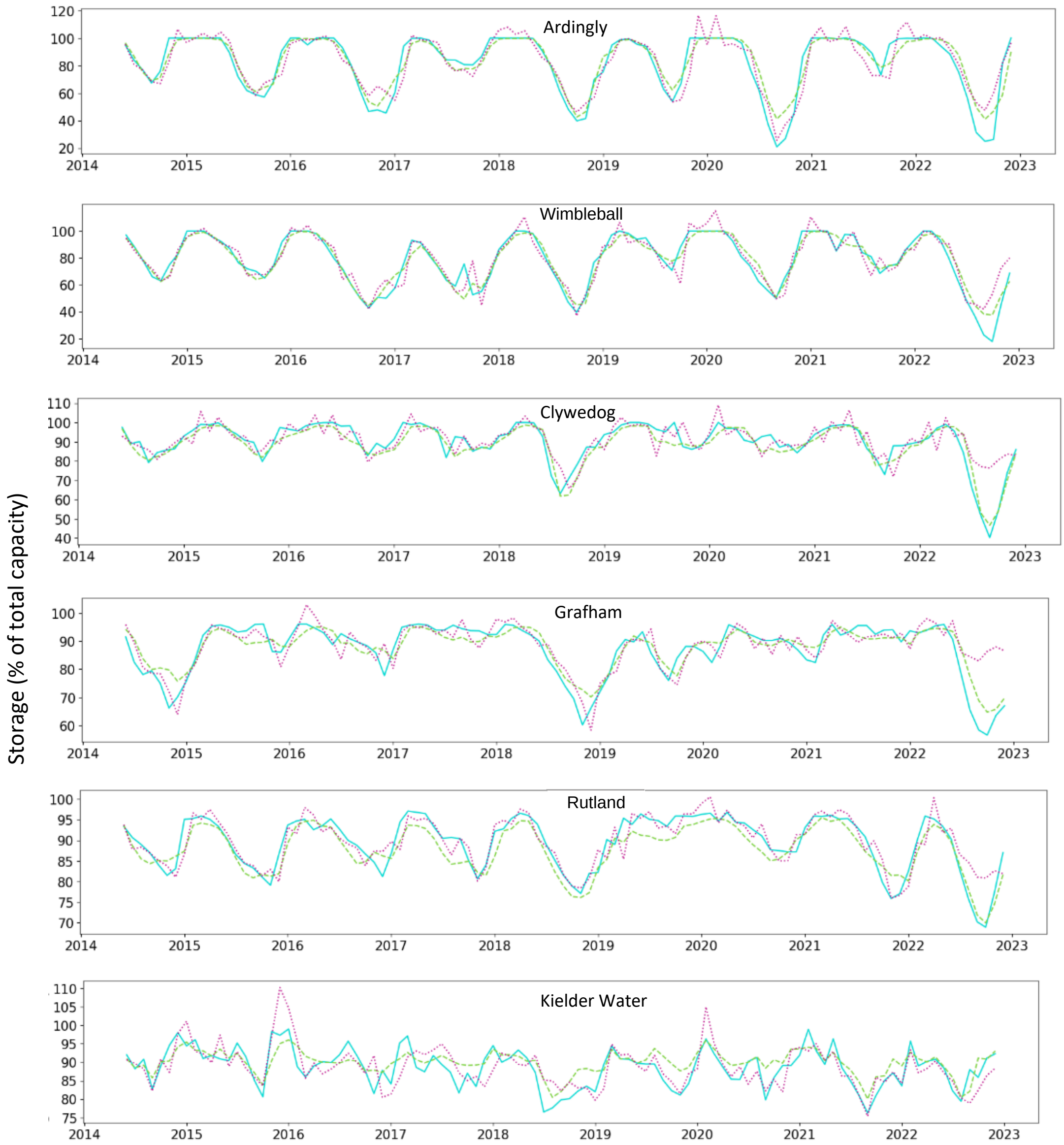

Figure 7: Reservoir storage: observed (solid blue); RF model (dashed green); ARIMA model (dotted pink), 1 month ahead predictions.

4.2 Forecasting

The RF model for Wimbleball was driven with climate forecasts using the historical analogue method employed in the UK hydrological forecasting to produce monthly and seasonal storage forecasts (UK Centre for Ecology & Hydrology, 2023). The seasonal climate forecasts consist of 10 resamples of 51 realizations, resulting in an ensemble of 510 climate forecasts that are each deemed equally likely. The seasonal climate forecasts for 2015 over the Wimbleball catchment are presented in Figure 8, with the distribution of the ensemble members represented by a violin plot and actual values shown as horizontal dashes. The reservoir storage forecast by the RF model is shown in Figure 9, against the historic storage percentile bands as used in the UK Water Resources Portal (Barker *et al.*, 2022).

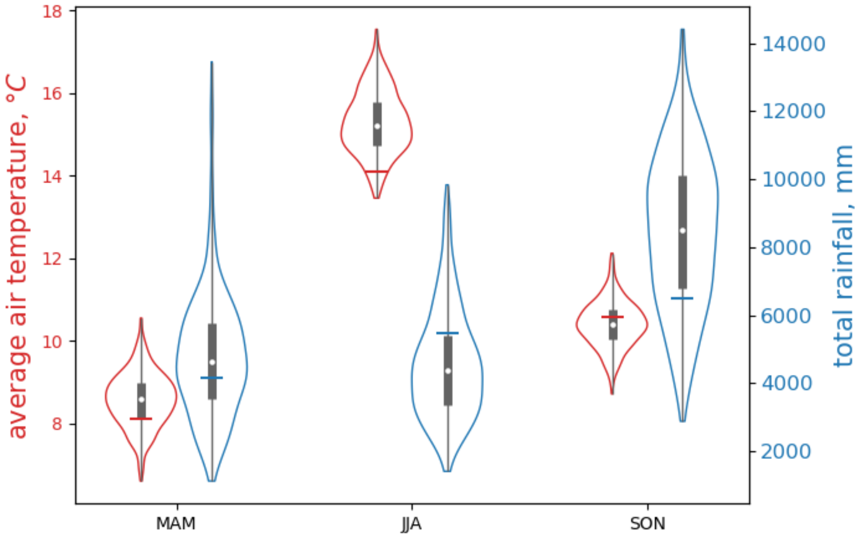

Figure 8: Seasonal climate forecasts over the Wimbleball catchment for 2015 (MAM: March April May; JJA: June July August; SON: September October November), with average air temperature in red and total rainfall summed over the catchment in blue. Actual values for average air temperature and total rainfall over the catchment are shown by horizontal dashes.

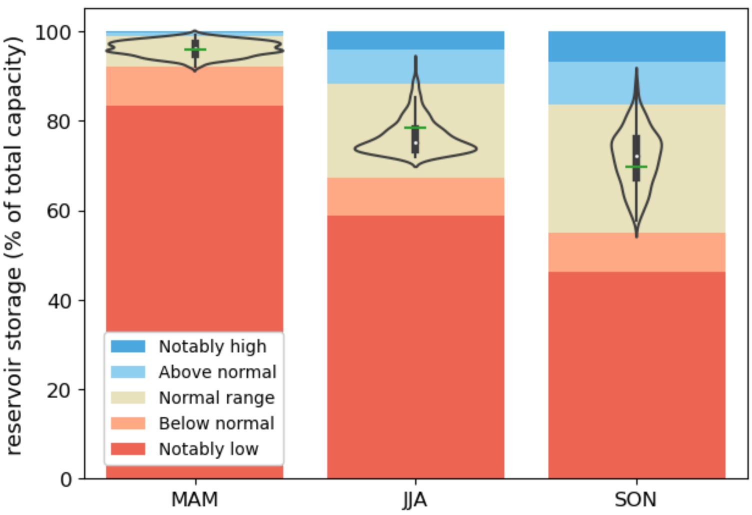

Figure 9: Seasonal forecasts of reservoir storage at Wimbleball for 2015 (MAM: March April May; JJA: June July August; SON: September October November), plotted above the historic storage percentile bands. Actual storage values are shown by horizontal dashes.

From Figure 9, it can be seen that the RF model for Wimbleball provides a good seasonal storage forecast for 2015, with the observed values falling within the 25th

and 75th forecast percentiles. Comparing Figure 9 and Figure 8, it is interesting to note that a high variability in climate forecasts does not always translate to a high variability in storage forecasts. For instance, the MAM storage forecast has low variability while the rainfall and air temperature have high variability. This is likely due to the reservoir being close to maximum capacity in February which limits the range of future forecasts in the MAM season.

The monthly climate forecast ensemble consists of 140 members, 10 resamples of 14 realizations, with results for total rainfall and average air temperature over the Wimbleball catchment for 2015/2016 shown in Figure 10. For this period the precipitation forecasts are relatively accurate over the Wimbleball catchment: actual precipitation always falls within the ensemble of forecasts and within the 25th – 75th percentiles for 5 out of 9 months. Temperature forecasts are reasonable for most of the months, but fail to accurately forecast a particularly warm November and December in 2015.

The monthly reservoir storage forecast by the RF model is shown in Figure 11, against the historic storage percentile bands as used in the UK Water Resources Portal (Barker *et al.*, 2022). The monthly storage forecasts are not as accurate as the seasonal ones for the same period: although the actual storage values are generally within the ensemble storage forecasts, they fall outside of the 25th – 75th percentiles for 7 out of the 9 months. This analysis should be extended to other reservoirs and longer time periods for a more robust assessment of model forecast skill, but is currently somewhat limited by the availability of climate hindcast data, which only extends to May 2016.

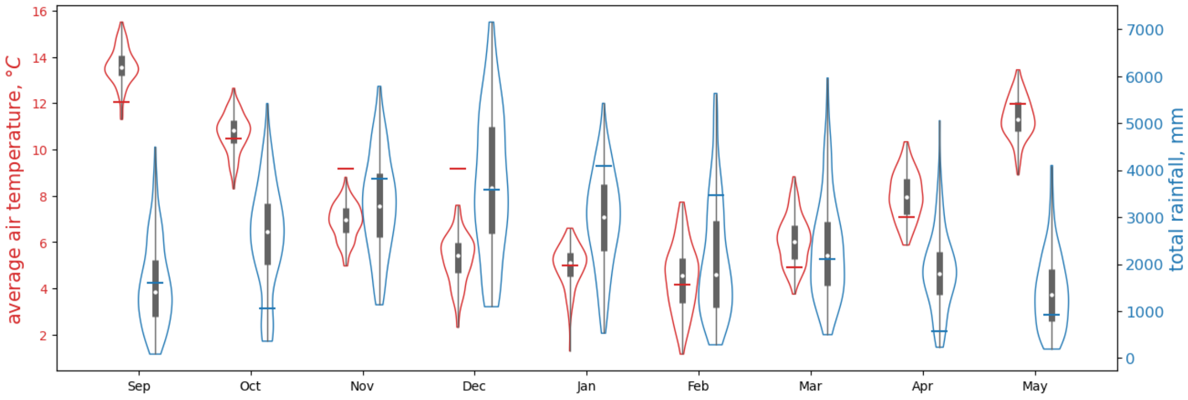

Figure 10: Monthly climate forecasts over the Wimbleball catchment for 2015/2016, with average air temperature in red and total rainfall summed over the catchment in blue. Actual values for average air temperature and total rainfall over the catchment are shown by horizontal dashes.

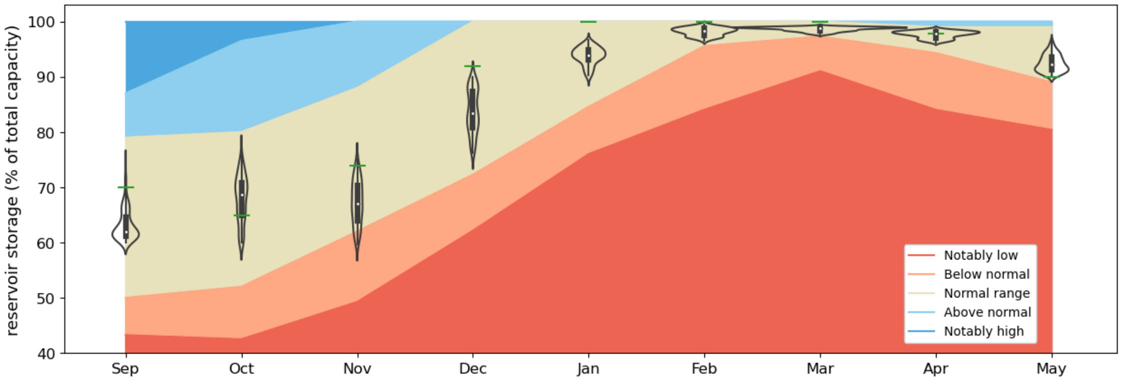

Figure 11: Monthly forecasts of reservoir storage at Wimbleball for 2015/2016, plotted above the historic storage percentile bands. Actual storage values are shown by horizontal dashes.

Summary: forecasting

Initial results suggest that RF methods can be successfully employed to predict monthly reservoir storage in many UK reservoirs, with good model performance at 5 out of the 6 case study reservoirs, and reasonable predictive skill extending even up to a 6 month lead time in 3 of the reservoirs. However, one limitation of data-driven methods is that they are less accurate when predicting out of the range seen in the training period, and this was evident in the poor model performance seen in the extreme low-storage periods. Future steps include the creation of a universal model, combining all reservoir storage time series along with static catchment characteristics, which should improve the model performance in these extreme low-storage periods.

These models have been combined with forecast climate data to produce monthly and seasonal storage forecasts in one example reservoir. These results look promising, and can be extended to the universal RF model.

5. Conclusions

- Three different status metrics were calculated for the Wimbleball reservoir to assess their suitability for outputs such as the UK Hydrological Summary and Outlooks, and the WMO HydroSOS. These metrics were displayed via graphical, tabular and spatial formats.
- The suitability of these metrics ultimately depends upon the needs of the end user e.g. whether they are interested in the current % of storage capacity, or identifying how current storage values compare to what is considered to be a 'normal' range for that particular month, season etc.
- Although the metrics have only been applied to current reservoir status here, they could also be used in the context of monthly to seasonal forecasting where deemed appropriate.
- Random Forest (RF) type models were applied to individual reservoirs: trained on monthly historic storage data, catchment temperature and rainfall, and used to predict reservoir storage at 1, 3, and 6 month lead time.
- These RF models showed model skill for 5 out of 6 of the selected reservoirs at 1 month lead time, and 3 out of 6 reservoirs at 6 month lead time.
- One model (at Wimbleball reservoir) was driven with monthly and seasonal (3 month) climate forecast ensembles, and showed promising results for reservoir storage forecasting.
- Model skill might be improved by combining storage data for multiple reservoirs with static catchment characteristics to create a "global" storage model.

6. Bibliography

Ahmad, S. K. and Hossain, F. (2019) A web-based decision support system for smart dam operations using weather forecasts. *Journal of Hydroinformatics*. 21 (5), pp. 687–707.

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

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B. and Lettenmaier, D. P. (2016) Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water resources research*. 52 (6), pp. 4209–4225.

Avisse, N., Tilmant, A., Müller, M. F. and Zhang, H. (2017) Monitoring small reservoirs' storage with satellite remote sensing in inaccessible areas. *Hydrology and Earth System Sciences*. 21 (12), pp. 6445–6459.

Barker, L. J., Fry, M., Hannaford, J., Nash, G., Tanguy, M. and Swain, O. (2022) Dynamic High Resolution Hydrological Status Monitoring in Real-Time: The UK Water Resources Portal. *Frontiers in Environmental Science*. 10.

Barker, L. J., Hannaford, J., Chiverton, A. and Svensson, C. (2015) *From meteorological to hydrological drought using standardised indicators*.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J. (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*. 4 (3), pp. 677–699.

Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., von Bloh, W. and Gerten, D. (2011) Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water resources research*. 47 (3).

Biswas, N. K., Hossain, F., Bonnema, M., Lee, H. and Chishtie, F. (2021) Towards a global Reservoir Assessment Tool for predicting hydrologic impacts and operating patterns of existing and planned reservoirs. *Environmental Modelling & Software*. 140p. 105043.

Boongaling, C. G. K., Faustino-Eslava, D. V. and Lansigan, F. P. (2018) Modeling land use change impacts on hydrology and the use of landscape metrics as tools for watershed management: The case of an ungauged catchment in the Philippines. *Land Use Policy*. 72pp. 116–128.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L. and Wada, Y. (2019) Development of the Community Water Model (CWatM v1.04). A high-resolution hydrological model for global and regional assessment of integrated water resources management. *Geoscientific Model Development Discussions*. 13 (7), pp. 3267–3298.

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., Pekel, J.-F. and Cottam, A. (2019) A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*. 23 (2), pp. 669–690.

Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Moreno-Saavedra, L. M., Morales-Díaz, B., Sanz-Justo, J., Gutiérrez, P. A. and Salcedo-Sanz, S. (2020) Analysis and Prediction of Dammed Water Level in a Hydropower Reservoir Using Machine Learning and Persistence-Based Techniques. *Water*. 12 (6), p. 1528.

Chinnasamy, P., Maheshwari, B. and Prathapar, S. A. (2018) Adaptation of Standardised Precipitation Index for understanding watertable fluctuations and groundwater resilience in hard-rock areas of India. *Environmental earth sciences*. 77 (15), p. 562.

Choong, S.-M. and El-Shafie, A. (2015) State-of-the-Art for Modelling Reservoir Inflows and Management Optimization. *Water Resources Management*. 29 (4), pp. 1267–1282.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C. and Cox, P. M. (2011) The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*. 4 (3), pp. 701–722.

Coerver, H. M., Rutten, M. M. and van de Giesen, N. C. (2018) Deduction of reservoir operating rules for application in global hydrological models. *Hydrology and Earth System Sciences*. 22 (1), pp. 831–851.

Cojan, I., Fouché, O., Lopéz, S. and Rivoirard, J. (2005) “Process-based Reservoir Modelling in the Example of Meandering Channel,” in Oy Leuangthong & Clayton V. Deutsch (eds.) *Geostatistics Banff 2004*. Quantitative geology and geostatistics. Dordrecht: Springer Netherlands. pp. 611–619.

Crétaux, J. F., Abarca-del-Río, R., Bergé-Nguyen, M., Arsen, A., Drolon, V., Clos, G. and Maisongrande, P. (2016) Lake Volume Monitoring from Space. *Surveys in Geophysics*. 37 (2), pp. 269–305.

Criminisi, A. (2011) Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Foundations and Trends® in Computer Graphics and Vision*. 7 (2-3), pp. 81–227.

Degenne, A., Bourgin, F., Perrin, C. and Andréassian, V. (2024) *Towards a better understanding of the hybrid modelling methodology for streamflow prediction*.

Döll, P., Kaspar, F. and Lehner, B. (2003) A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of hydrology*. 270 (1-2), pp. 105–134

Donchyts, G., Winsemius, H., Baart, F., Dahm, R., Schellekens, J., Gorelick, N., Iceland, C. and Schmeier, S. (2022) High-resolution surface water dynamics in Earth's small and medium-sized reservoirs. *Scientific Reports*. 12 (1), p. 13776.

Droppers, B., Franssen, W. H. P., van Vliet, M. T. H., Nijssen, B. and Ludwig, F. (2020) Simulating human impacts on global water resources using VIC-5. *Geoscientific Model Development*. 13 (10), pp. 5029–5052.

Ehsani, N., Fekete, B. M., Vörösmarty, C. J. and Tessler, Z. D. (2016) A neural network based general reservoir operation scheme. *Stochastic environmental research and risk assessment : research journal*. 30 (4), pp. 1151–1166.

Environment Agency (2024) *Water situation: national monthly reports for England 2024* [Online]. Available from: https://www.gov.uk/government/publications/water-situation-national-monthly-reports-for-england-2024 (Accessed 11 April 2024).

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M. and Tarboton, D. (2016) An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of hydrology*. 537pp. 45–60.

Fernando, D. N., Zhu, J. Z. and Negusse, S. (2017) *Framework for the application of seasonal rainfall forecasts' ' for reservoir storage forecasts: proof of concept for three reservoirs in the Brazos River Basin*.

Ficchì, A., Raso, L., Dorchies, D., Pianosi, F., Malaterre, P. O., Van Overloop, P. J. and Jay-Allemand, M. (2016) Optimal operation of the multireservoir system in the seine river basin using deterministic and ensemble forecasts. *Journal of Water Resources Planning and Management*. 142 (1), p. 05015005.

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

García-Feal, O., González-Cao, J., Fernández-Nóvoa, D., Astray Dopazo, G. and Gómez-Gesteira, M. (2022) Comparison of machine learning techniques' ' for reservoir outflow forecasting. *Natural Hazards and Earth System Science*. 22 (12), pp. 3859–3874.

Gleick, P. H. (2003) Global freshwater resources: soft-path solutions for the 21st century. *Science*. 302 (5650), pp. 1524–1528.

Gleick, P. H., Cooley, H., Famiglietti, J. S., Lettenmaier, D. P., Oki, T., Vörösmarty, C. J. and Wood, E. F. (2013) *Improving understanding of the global hydrologic cycle, in: Climate science for serving society*. p.pp. 151–184.

Gronewold, A. D., Clites, A. H., Hunter, T. S. and Stow, C. A. (2011) An appraisal of the Great Lakes advanced hydrologic prediction system. *Journal of Great Lakes research*.

Gusyev, M. A. (2015) "*Drought assessment in the Pampanga River basin, the Philippines – Part 1: Characterizing a role of dams in historical droughts with standardized indices*," in 2015 Queensland, Australia: MODSIM.

Gusyev, M., Hasegawa, A., Magome, J., Sanchez, P., Sugiura, A., Umino, H., Sawano, H., Tokunaga, Y., International Centre for Water Hazard and Risk Management (ICHARM), Public Works Research Institute (PWRI), 1-6 Minamihara, Tsukuba, Ibaraki 305-8516, Japan, National Graduate Institute for Policy Studies (GRIPS), Tokyo, Japan, International Research Centre for River Basin Environment (ICRE), University of Yamanashi, Kofu, Japan and UNESCO Office Jakarta, Jakarta, Indonesia (2016) Evaluation of Water Cycle Components with Standardized Indices Under Climate Change in the Pampanga, Solo and Chao Phraya Basins. *Journal of Disaster Research*. 11 (6), pp. 1091–1102.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z. D., Wada, Y. and Wisser, D. (2014) Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences of the United States of America*. 111 (9), pp. 3251–3256.

Hanasaki, N., Kanae, S. and Oki, T. (2006) A reservoir operation scheme for global river routing models. *Journal of hydrology*. 327 (1-2), pp. 22–41.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. and Tanaka, K. (2008) An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*. 12 (4), pp. 1007–1025.

Hong, J., Lee, S., Bae, J. H., Lee, J., Park, W. J., Lee, D., Kim, J. and Lim, K. J. (2020) Development and evaluation of the combined machine learning models for the prediction of dam inflow. *Water*. 12 (10), p. 2927.

Hughes, D., Birkinshaw, S. and Parkin, G. (2021) A method to include reservoir operations in catchment hydrological models using SHETRAN. *Environmental Modelling & Software*. 138p. 104980.

Hunt, K. M. R., Matthews, G. R., Pappenberger, F. and Prudhomme, C. (2022) Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. *Hydrology and Earth System Sciences*. 26 (21), pp. 5449–5472.

Ibañez, S. C., Dajac, C. V. G., Liponhay, M. P., Legara, E. F. T., Esteban, J. M. H. and Monterola, C. P. (2021) Forecasting reservoir water levels using deep neural networks: A case study of angat dam in the philippines. *Water*. 14 (1), p. 34.

Kim, J., Read, L., Johnson, L. E., Gochis, D., Cifelli, R. and Han, H. (2020) An experiment on reservoir representation schemes to improve hydrologic prediction: coupling the national water model with the HEC-ResSim. *Hydrological Sciences Journal*. 65 (10), pp. 1652–1666.

Kim, S.-J., Bae, S.-J., Lee, S.-J. and Jang, M.-W. (2022) Monthly agricultural reservoir storage forecasting using machine learning. *Atmosphere*. 13 (11), p. 1887.

Kraft, B., Jung, M., Körner, M., Koirala, S. and Reichstein, M. (2022) Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences*. 26 (6), pp. 1579–1614.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S. and Nearing, G. (2019) Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*. 23 (12), pp. 5089–5110.

Li, B., Friedman, J., Olshen, R. and Stone, C. (1984) Classification and regression trees (CART). *Biometrics*. 40 (3), pp. 358–361.

McKee, T., Doesken, N. and Kleist, J. (1993) "*The relationship of drought frequency and duration to timescales*," in 17 January 1993 Eighth Conference of Applied Climatology, Colorado. pp. 179–183.

Met Office, Hollis, McCarthy, Kendon, Legg and Simpson (2018) *HadUK-Grid gridded and regional average climate observations for the UK*. [Online]. Available from: http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb (Accessed 5 March 2024). [Online]. Available from: http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb (Accessed 5 March 2024).

Nazemi, A. and Wheater, H. S. (2015) On inclusion of water resource management in Earth system models – Part 1: Problem definition and representation of water demand. *Hydrology and Earth System Sciences*. 19 (1), pp. 33–61.

Ng, K. W., Huang, Y. F., Koo, C. H., Chong, K. L., El-Shafie, A. and Najah Ahmed, A. (2023) A review of hybrid deep learning applications for streamflow forecasting. *Journal of hydrology*. 625p. 130141.

Nilsson, C., Reidy, C. A., Dynesius, M. and Revenga, C. (2005) Fragmentation and flow regulation of the world's large river systems. *Science*. 308 (5720), pp. 405–408.

NRFA (2022) *Monthly Hydrological Summaries | National River Flow Archive* [Online]. Available from: https://nrfa.ceh.ac.uk/monthly-hydrological-summary-uk (Accessed 18 March 2022).

NRFA (2024) *Monthly Hydrological Summaries | National River Flow Archive* [Online]. Available from: https://nrfa.ceh.ac.uk/monthly-hydrological-summary-uk (Accessed 6 March 2024).

Özdoğan-Sarıkoç, G., Sarıkoç, M., Celik, M. and Dadaser-Celik, F. (2023) Reservoir volume forecasting using artificial intelligence-based models: Artificial Neural Networks, Support Vector Regression, and Long Short-Term Memory. *Journal of hydrology*. 616p. 128766.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.

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

Raghavendra. N, S. and Deka, P. C. (2014) Support vector machine applications in the field of hydrology: A review. *Applied soft computing*. 19pp. 372–386.

Rajesh, M., Anishka, S., Viksit, P. S., Arohi, S. and Rehana, S. (2023) Improving Short-range Reservoir Inflow Forecasts with Machine Learning Model Combination. *Water Resources Management*. 37 (1), pp. 75–90.

Rickards, N. J. and Baron, H. E. (2022) *Global water tools: Water body storage, status and forecasting*. p.p. 40.

Rickards, N. J., Chengot, R., Baron, H. E. and Fry, M. (2023) *Copernicus In Situ: Review of global in situ data for lakes and reservoirs* . p.p. 10.

Salwey, S., Coxon, G., Pianosi, F., Singer, M. B. and Hutton, C. (2023) National-scale detection of reservoir impacts through hydrological signatures. *Water resources research*. 59 (5), .

Sapitang, M., M. Ridwan, W., Faizal Kushiar, K., Najah Ahmed, A. and El-Shafie, A. (2020) Machine learning application in reservoir water level forecasting for sustainable hydropower generation strategy. *Sustainability*. 12 (15), p. 6121.

Steyaert, J. C. and Condon, L. E. (2023) *Synthesis of Historical Reservoir Operations from 1980 – 2020 for the Evaluation of Reservoir Representation in Large Scale Hydrologic Models*.

Svoboda, M. (2012) *The Standardized Precipitation Index User Guide* Deborah A Wood (ed.).

Tiwari, A. D. and Mishra, V. (2019) Prediction of reservoir storage anomalies in India. *Journal of Geophysical Research: Atmospheres*.

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

Turner, S. W. D., Doering, K. and Voisin, N. (2020) Data-driven reservoir simulation in a large-scale hydrological and water resource model. *Water resources research*. 56 (10), .

UK Centre for Ecology & Hydrology (2023) *UK Hydrological Outlook* [Online]. Available from: https://www.ceh.ac.uk/news-and-media/blogs/uk-hydrological-outlook-upgraded-new-forecasts-based-historic-weather (Accessed 6 March 2024).

UK Met Office (2022) *UK actual and anomaly maps - Met Office* [Online]. Available from: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-actual-and-anomaly-maps (Accessed 24 March 2022).

Valipour, M., Banihabib, M. E. and Behbahani, S. M. R. (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of hydrology*. 476pp. 433–441.

Vieira Valadão, L., Ennes Cicerelli, R., de Almeida, T., Barbosa Curto Ma, J. and Garnier, J. (2021) Reservoir metrics estimated by remote sensors based on the Google Earth Engine platform. *Remote Sensing Applications: Society and Environment*. 24p. 100652.

Wada, Y., van Beek, L. P. H. and Bierkens, M. F. P. (2012) Nonsustainable groundwater sustaining irrigation: A global assessment. *Water resources research*. 48.

Wang, F., Wang, L., Zhou, H., Saavedra Valeriano, O. C., Koike, T. and Li, W. (2012) Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions. *Water resources research*. 48 (7).

Wang, Q. and Wang, S. (2020) Machine Learning-Based Water Level Prediction in Lake Erie. *Water*. 12 (10), p. 2654.

World Meteorological Organisation (2012) *Standardized Precipitation Index' ' User Guide*.

World Meteorological Organization (2020) *Global Hydrological Status and Outlook System (HydroSOS)* [Online]. Available from: https://community.wmo.int/activity-areas/global-hydrological-status-and-outlook-system-hydrosos (Accessed 4 April 2022).

Xu, S., Chen, Y., Xing, L. and Li, C. (2021) Baipenzhu reservoir inflow flood forecasting based on a distributed hydrological model. *Water*. 13 (3), p. 272.

Yang, T., Asanjan, A. A., Welles, E., Gao, X., Sorooshian, S. and Liu, X. (2017) Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water resources research*. 53 (4), pp. 2786–2812.

Yang, T., Gao, X., Sorooshian, S. and Li, X. (2016) Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water resources research*. 52 (3), pp. 1626–1651.

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G. and Wheater, H. (2019) Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrology and Earth System Sciences*. 23 (9), pp. 3735–3764.

Zarei, M., Bozorg-Haddad, O., Baghban, S., Delpasand, M., Goharian, E. and Loáiciga, H. A. (2021) Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages. *Scientific Reports*. 11 (1), p. 24295.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X. and Zhuang, J. (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of hydrology*. 565pp. 720–736.

Zhao, G., Gao, H., Naz, B. S., Kao, S.-C. and Voisin, N. (2016) Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in water resources*. 98pp. 16–31.

Zhu, S., Hrnjica, B., Ptak, M., Choiński, A. and Sivakumar, B. (2020) Forecasting of water level in multiple temperate lakes using machine learning models. *Journal of hydrology*. 585p. 124819.

7. Appendix

This appendix includes additional details on the Random Forest models built for reservoir storage prediction and forecasting in Section 4.

Table A 1: Summary of the machine learning models that were explored for each reservoir.

Machine learning Models used	Details	Advantages	Disadvantages	Hyper parameters
KNeighbors Regressor	Instance-based regression algorithm where predictions are made based on the k-nearest neighbors	-Simple to understand and implement - Robust to noisy training data	-Computationally expensive for large datasets - sensitive to the choice of optimal number of neighbors	'n_neighbors'
Decision Tree Regressor	Builds a regression model in the form of a tree structure, where each internal node represents a decision based on a feature	Intuitive, simple to use, ML algorithm that can detect highly non-linear relationships in datasets. Decision trees are fast to train, and easy to explain.	Prone to overfitting, sensitive to small variations in data	'max_depth' 'min_samples_split' 'ccp_alpha'
Extra Tree Regressor	Similar to Decision Trees, but it randomly selects feature splits, leading to a broader exploration of the feature space	Less prone to overfitting compared to Decision Trees	May not work well with noisy data	'max_depth' 'min_samples_split' 'min_samples_leaf' 'max_features'
Support Vector Regression	Support Vector Regression uses support vector machines to perform regression	- Performs well in Higher dimension, - Best algorithm when classes are separable - Outliers have less impact	- Slow for large datasets - Poor performance with overlapped classes	'C' 'epsilon' 'kernel' 'gamma'

			- Sensitive to choice of kernel parameters.	
Ada Boost Regressor	Boosting ensemble method that combines multiple weak learners to create a strong learner	-Resistant to overfitting,- -Handles outliers well	-Sensitive to noisy data, - can be slow to train	'n_estimators' 'learning_rate' 'base_estimators' such as max_depth or min_samples_split
Bagging Regressor	Ensemble averaging method that builds multiple base models on bootstrapped samples and averages their predictions	-Works best with strong and complex models	- Doesn't improve performance if the base model is biased	'max_samples' 'max_features' 'bootstrap'/ 'bootstrap_features'
Random Forest Regressor	Ensemble learning method that constructs a multitude of decision trees and averages their predictions	-Robustness to overfitting -Easy to use -The default parameters often give very good results -Parameter tuning is straightforward.	- slower to train and score -difficult to predict complex, large or sparse datasets	'n_estimators' 'min_samples_split' ' 'max_features'
Extra Trees Regressor	Ensemble learning method similar to RandomForest, but with random splits for each decision tree	Reduces variance and overfitting, faster training	May require more trees to achieve optimal performance	'n_estimators' 'max_depth' 'min_samples_split' 'min_samples_leaf' 'max_features' 'bootstrap'
Gradient Boosting Regressor	Boosting ensemble method that that tries to combine many weak learners into a stronger whole	-currently best-in-class in terms of performance for structured ML problems.	-a lot of parameters to consider -little worse in overfitting.	'n_estimators' 'learning_rate' 'max_depth' 'gamma' 'subsample'

		- handle sparse/sparse data much better		'max_features'

Table A 2: Features used for predicting reservoir storage at time t.

Features	Description
rainfall	Monthly rainfall over the catchment for current month
tas	Monthly mean air temperature over the catchment for current month
percent_full{i} for i=1 to 12	Reservoir storage as a percentage of total capacity at time t-i
rainfall{i} for i=1 to 6	Monthly rainfall over the catchment at time t-i
tas {i} for i=1 to 6	Monthly mean air temperature over the catchment at time t-i
mean_month	Long-term mean reservoir storage for each month
rf{quart, half, yr}	Mean of catchment rainfall over the last {3,6,12} months (including current month)
tas{quart, half, yr}	Mean of average air temperature over the last {3,6,12} months (including current month)
datesin, datecos	Cyclical calendar variables

Table A 3: Reservoirs and model parameters used for Random Forest

Reservoir	Importance Threshold	Parameters
Ardingly	0.006	'percent_full1', 'percent_full2', 'mean_month', 'datesin', 'rainfall', 'tas', 'rainfall1', 'tas1', 'tas2', 'tas3', 'tas6', 'rfquart', 'rfhalf', 'rfyr', 'tasquart', 'tashalf'
Wimbleball	0.004	'percent_full1', 'percent_full2', 'percent_full3', 'datesin', 'datecos', 'rainfall', 'tas', 'rainfall1', 'tas1', 'tas2', 'tas5', 'tas6', 'rfquart', 'rfhalf', 'rfyr', 'tasquart', 'tasyr'
Clywedog	0.015	'percent_full1', 'percent_full2', 'mean_month', 'datesin', 'datecos', 'rainfall', 'tas', 'tas2', 'tas3', 'tas5', 'rfquart', 'rfhalf', 'tasquart', 'tashalf'

Grafham	0.007	'percent_full1', 'percent_full2', 'percent_full3', 'percent_full10', 'percent_full11', 'mean_month', 'datesin', 'tas', 'tas1', 'tas2', 'tas6', 'rfquart', 'rfhalf', 'tasquart', 'tashalf', 'tasyr'
Rutland	0.008	'percent_full1', 'percent_full2', 'mean_month', 'datesin', 'tas', 'tas1', 'tas2', 'tas3', 'tas6', 'rfquart', 'rfhalf', 'rfyr', 'tasquart', 'tashalf'
Kielder Water	0.011	'percent_full1', 'percent_full2', 'percent_full3', 'percent_full12', 'mean_month', 'datesin', 'rainfall', 'rainfall1', 'rainfall5', 'tas1', 'tas2', 'tas6', 'rfquart', 'rfhalf', 'rfyr', 'tasquart', 'tashalf', 'tasyr'

Contact

UK Centre for Ecology & Hydrology
Maclean Building, Benson Lane
Crowmarsh Gifford
Wallingford, Oxfordshire
OX10 8BB

t: +44 (0)1491 838800
e: NC-international@ceh.ac.uk

UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW

t: +44 (0)1248 374500
e: NC-international@ceh.ac.uk

UK Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB

t: +44 (0)131 4454343
e: NC-international@ceh.ac.uk

UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP

t: +44 (0)1524 595800
e: NC-international@ceh.ac.uk

UKRI Natural Environment Research Council

Our National Capability for Global Challenges programme is funded by the UKRI Natural Environment Research Council.