Hybrid forecasting: combining dynamical predictions with data-driven models

Louise J. Slater[1], Louise Arnal[2], Marie-Amelie Boucher[3], Annie Y.-Y. Chang[4,5], Simon Moulds[1], Conor Murphy[6], Grey Nearing[7], Guy Shalev[8], Chaopeng Shen[9], Linda Speight[1], Gabriele Villarini[10], Robert L. Wilby[11], Andrew Wood[12], and Massimiliano Zappa[4]

[1]School of Geography and the Environment, University of Oxford, Oxford, UK
[2]University of Saskatchewan, Centre for Hydrology, Canmore, Canada
[3]Université de Sherbrooke, Canada
[4]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[5]ETH, Zurich, Switzerland
[6]Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Kildare, Ireland
[7]Google Research, Mountain View, CA, USA
[8]Google Research, Tel Aviv, Israel
[9]Civil and Environmental Engineering, The Pennsylvania State University, State College, PA 16801, USA
[10]IIHR–Hydroscience and Engineering, University of Iowa, Iowa, USA
[11]Geography and Environment, Loughborough University, Loughborough, UK
[12]National Center for Atmospheric Research, Climate and Global Dynamics, Boulder, CO, USA

Correspondence: Louise J. Slater (louise.slater@ouce.ox.ac.uk)

Abstract. Hybrid hydroclimatic forecasting systems employ data-driven (statistical or machine learning) methods to harness and integrate a broad variety of predictions from dynamical, physics-based models – such as numerical weather prediction, climate, land, hydrology and Earth System models – into a final prediction product. They are recognised as a promising way of enhancing prediction skill of meteorological and hydroclimatic variables and events, including rainfall, temperature, streamflow, floods, droughts, tropical cyclones, or atmospheric rivers. Hybrid forecasting methods are now receiving growing attention due to advances in weather and climate prediction systems at sub-seasonal to decadal scales, a better appreciation of the strengths of machine learning, as well as expanding access to computational resources and methods. Such systems are attractive because they may avoid the need to run a computationally-expensive offline land model, can minimize the effect of biases that exist within dynamical outputs without explicit bias correction and downscaling, benefit from the strengths of machine learning, and can learn from large datasets, while combining different sources of predictability with varying time-horizons. Here we review recent developments in hybrid hydroclimatic forecasting and outline key challenges and opportunities for further research. These include obtaining physically-explainable results, assimilating human influences from novel data sources, integrating new ensemble techniques to improve predictive skill, creating seamless prediction schemes that merge short to long lead times, incorporating modelled initial land surface and ocean/ice conditions, acknowledging spatial variability in landscape and atmospheric forcing, and increasing the operational uptake of hybrid prediction schemes.

1 Introduction: Defining hybrid forecasting and prediction

This review addresses the growing popularity of hybrid forecasting, an approach that seeks to enhance the predictability of hydroclimatic variables by merging predictions from 'dynamical' physics-based weather or climate simulation models with data-driven models. Dynamical models represent temporal changes in system properties by using numerical modelling to solve dynamical physical processes. Data-driven models include empirical, statistical and machine learning (ML) methods, and can range from simple linear regression to deep neural networks. Recognising that dynamical and empirical models have different strengths, hybrid prediction reflects the deliberate fusing of the two.

While challenging to identify distinct categories, given the flexibility and diversity of hybrid methods, three principal types of hybrid model structure may be discerned (Figure 1; Table 1). (i) Statistical-dynamical models typically drive a statistical or ML model (data-driven) with dynamical weather or climate model outputs from numerical weather prediction (NWP) models or Earth System Models (ESMs). The statistical-dynamical structure is the most common type of hybrid model in the literature (Table 2). (ii) Serial models combine data-driven and dynamical models sequentially, and may include additional types of models such as a hydrological model. (iii) Coupled or parallel approaches combine data-driven and dynamical models in parallel. The coupled approach is more commonly employed in operational settings, where ML is increasingly being used to upgrade components within existing modelling schemes. We do not provide a prescriptive definition of hybrid forecasting as it exists along a continuum from loosely to 'fully' hybrid (e.g. AghaKouchak et al., 2022), and may include a wide range of models and 'big data', such as Earth Observations (EO).

Table 1. Examples of different hybrid model structures.

Name	Description
(i) Statistical-dynamical	Statistical-dynamical hybrid models consist of driving or conditioning a data-driven model with dynamical weather, climate, or Earth System Model (ESM) predictions (e.g. Vecchi et al., 2011; Slater and Villarini, 2018). Both expressions 'statistical-dynamical' and 'dynamical-statistical' are used depending on the focus of the research or the field of study. This approach is also referred to as 'parameter informed' (e.g. Schlef et al., 2021) or 'physical–statistical' (e.g. AghaKouchak et al., 2022) prediction.
(ii) Serial	A serial structure combines the dynamical and data-driven models sequentially, and may include additional models such as a hydrological model. For instance, one could pre-/post-process the output of a dynamical model using a data-driven approach (e.g. Glahn and Lowry, 1972) and use those predictions as input to a conceptual or physics-based model. In Bennett et al. (2016), post-processed General Circulation Model (GCM) forecasts are used to force a monthly rainfall-runoff model. In Richardson et al. (2020), weather patterns are identified in an ensemble prediction system and subsequently used to forecast threshold exceedance probabilities of extreme precipitation and flooding.
(iii) Coupled or Parallel	In a coupled hybrid structure, the data-driven and dynamical model are combined in parallel. This may involve, for instance, replacing a component of a dynamical model with a data-driven model, e.g. to create a machine-learning corrected GCM (e.g. Watt-Meyer et al., 2021). Alternatively, it is possible to combine outputs from an ensemble of dynamical and statistical predictions run in parallel (e.g. Madadgar et al., 2016). A data-driven model may also be employed to combine dynamical predictions from both meteorological and hydrological models (e.g. Bogner et al., 2019).

Traditional workflows in which a physics-based or conceptual land/hydrology model generates the final forecast product are still the most commonly used operational forecasting systems worldwide. These may include 'physics-based' models, based on a spatially-distributed representation of known physical laws through mathematical equations and numerical solution

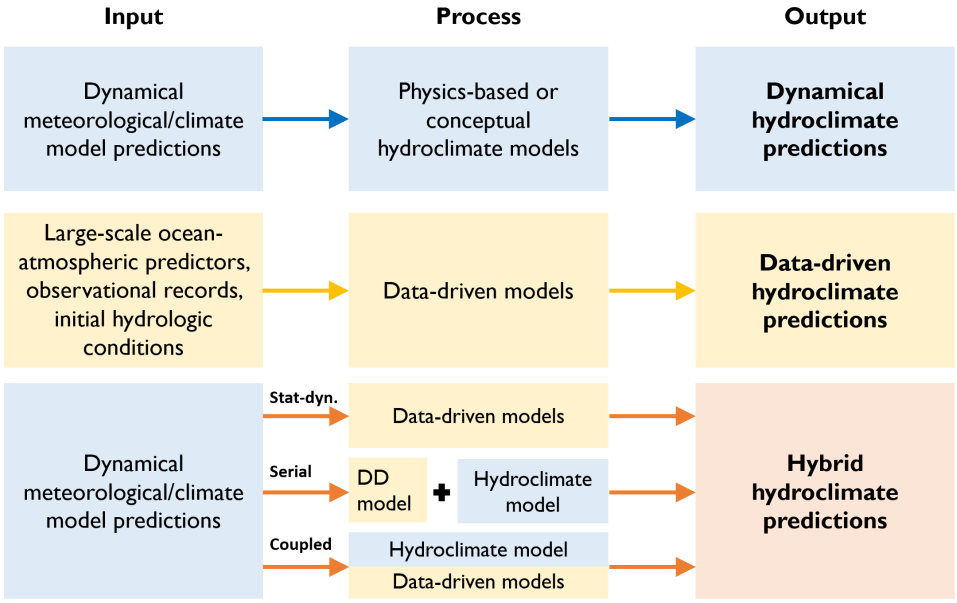

Figure 1. Defining hybrid hydroclimate forecasting and prediction. 'Hydroclimate' refers to a range of variables defined in the text, including streamflow. The top row indicates traditional dynamical hydroclimate predictions (blue); middle row is data-driven (DD) predictions (yellow) and bottom row represents hybrid predictions (red), which combine dynamical and data-driven approaches. In the last row, three examples of hybrid structure are shown from top to bottom: (i) Statistical-dynamical (Stat-dyn), (ii) Serial, and (iii) Coupled, as described in Table 1. The figure provides simple examples, but other schemes are possible, including for example a mix of observations and predictions in the left column.

(e.g. Freeze and Harlan, 1969), or 'conceptual' models, which simplify the representation of physical processes, often using empirical relationships (e.g. Nash and Sutcliffe, 1970). There is a long history of development and application of standalone dynamical land surface and catchment hydrology models of varying complexity (from conceptual to physically-explicit) for operational forecasting. Process-based hydrological modelling approaches may be either spatially distributed (gridded) or lumped (catchment-averaged). Examples include the hourly conceptual rainfall-runoff GR4H model used by the Bureau of Meteorology in Australia (Hapuarachchi et al., 2022); the conceptual reservoir-based HSAMI model implemented by Hydro-Québec (Bisson and Roberge, 1983); or the conceptual Sacramento Soil Moisture Accounting (SAC-SMA) model of the Community Hydrologic Prediction System of the U.S. National Weather Service (Burnash et al., 1973). In operational systems, the hydrological model is typically forced with NWP-based forecast meteorology, as in the case of the US National Water Model (NOAA, 2016) (see Zappa et al. (2008) for a report on science-driven operational application of several end-to-end ensemble hydrometeorological forecasting systems.) Outputs from coupled atmosphere-ocean-land GCMs may be used over longer time horizons, as is the case with the European and Global Flood Awareness Systems, EFAS and GloFAS (Alfieri et al., 2013; Thielen et al., 2009; Smith et al., 2016; Arnal et al., 2018; Emerton et al., 2018; Harrigan et al., 2020). These approaches are considered as more physically interpretable than 'black box' statistical methods. However, the large computational demand

and variable skill of many traditional forecasting approaches still persists (Arnal et al., 2018), and their calibration still requires substantial effort (Arheimer et al., 2020; Hirpa et al., 2018) relative to most data-driven models (see Section 3.4).

In contrast with traditional forecast workflows, data-driven prediction has historically relied more on observed data than dynamical climate model predictions, building empirical relationships between e.g. streamflow and precipitation (Garen, 1992), using time-lag relationships between upstream and downstream flow, or stochastic autoregression approaches like auto-regressive moving average models (Jain et al., 2018). In such data-driven models, the hydroclimatological predictands can be regressed on a range of covariates, such as observed precipitation/temperature records, static variables (e.g. elevation, slope, geology), initial hydrologic conditions, or large-scale predictors such as sea surface temperatures (SST), surface air temperature, geopotential height, meridional wind, sea ice extent, or modes of climate variability such as the El Niño-Southern Oscillation (ENSO) (e.g. Wilby et al., 2004; Dixon and Wilby, 2019; Mendoza et al., 2017; Meißner et al., 2017). Broadly speaking, the strength of statistical models lies in their simplicity, speed, ease-of-use, and comparable skill to dynamical methods when there are sufficient observations for model training. However, data-driven models are sometimes thought to be less able to extrapolate to extreme outlier values that have not been seen in the historical record (Milly et al., 2008; Frame et al., 2022a; Reichstein et al., 2019) or unable to reflect shifts in the relationship between the predictand and predictors. Others have raised the risk of artificial skill in cases where predictors are selected preferentially based on correlation with the predictand and not fully cross-validated (e.g. DelSole and Shukla, 2009). Data-driven models may also be difficult to optimize for multi-variate, high-dimensional output fields, which are simulated intrinsically by dynamical models. Recent studies focusing on more complex data-driven techniques such as deep learning have suggested that some of these limitations can be overcome, such as the extrapolation to extreme or unforeseen conditions (Frame et al., 2022a), to new (untrained) catchments (Kratzert et al., 2019a), and to poorly gauged large regions (Feng et al., 2021; Ma et al., 2021). Nevertheless, the inclusion of physical constraints could further elevate prediction robustness in data-sparse situations (Feng et al., 2022a). Research is required to understand the hydroclimatological conditions to which new ML and DL models are able to extrapolate from the training set, and their performance as they are extrapolated in space.

Hybrid forecasts benefit from combining the ability of physical models to predict and explain large-scale phenomena (i.e. through NWPs or climate model predictions) *with* the ability of data-driven models to efficiently estimate the characteristics of events from observed data and account for bias or anomalies in the data. Many current examples of hybrid prediction build on traditional forecast workflows by using an ML algorithm in sequence with or alongside a conceptual or physics-based hydrological model (World Meteorological Organization, 2021) (Figure 1). Some notable examples of operational hybrid prediction include the 'objective consensus' climate forecast (i.e. derived objectively from multiple models) at the US Climate Prediction Center, which uses ensemble regression (e.g. Unger et al., 2009) to combine multiple dynamical and statistical forecasts into one. The International Research Institute for Climate and Society (IRI) has a multi-model calibrated prediction based on three Subseasonal Experiment (SubX) models (Pegion et al., 2019). The UK Met Office uses a tool called 'Decider' which assigns medium-range precipitation forecast ensemble members to a set of 30 probabilistic weather patterns (Neal et al., 2016) and then feeds several downstream forecasting applications, such as for coastal flooding (Neal et al., 2018) and fluvial flooding (Richardson et al., 2020). Lastly, the Google flood forecasting model (https://sites.research.google/floods/) produces

operational, public-facing forecasts of water levels up to six days ahead (Nevo et al., 2022) using ML models forced with operational, real-time weather forecasts from the ECMWF Atmospheric Model high resolution 10-day forecast (ECMWF HRES) as inputs. Broadly speaking, many hydroclimate projection systems are now hybrid, as per the 'serial' definition in Table 1, because some kind of statistical processing is applied to generate a final information product from an ensemble of climate model outputs. Dynamical modelling centres often lack the resources or scope to tailor outputs to particular stakeholder needs (adding value with data-driven methods), leading to implementation of such processing by the end users themselves. These predictions are not always visible as 'hybrid' activity but are operational nonetheless. These examples show the general evolution of the field from traditional forecasting (Cohen et al., 2019) toward hybrid prediction.

The diversity of approaches for hybrid forecasting and prediction is evident from the sample of studies listed in Table 2. The scope of hybrid models can vary widely, encompassing different forecast units (e.g. hourly or seasonal mean forecasts), lead times (from the next hour to next decade, e.g. Ravuri et al., 2021; Neri et al., 2019), and geographical domains (from point to street-level, single river catchment through to global approaches). Hybrid models have been applied to predict a variety of hydrometeorological variables, including extreme heat and precipitation (Najafi et al., 2021; Miao et al., 2019; Ma et al., 2022), seasonal climate variables (Golian et al., 2022; Baker et al., 2020), tropical cyclones/hurricanes (Vecchi et al., 2011; Murakami et al., 2016; Kang and Elsner, 2020; Klotzbach et al., 2020), streamflow (Wood and Schaake, 2008; Mendoza et al., 2017; Rasouli et al., 2012; Duan et al., 2020), flooding (Slater and Villarini, 2018), drought (Madadgar et al., 2016; Wu et al., 2022), sea level (Khouakhi et al., 2019), and reservoir levels (Tian et al., 2021), over a range of timescales (Table 2). Certain other examples discussed in this review are not fully hybrid (e.g. ML models that are not driven by NWM/ESM predictions) but serve to illustrate the possibilities of future hybrid systems. Many types of data-driven models have been used (Tables 2-3), including simple regression methods, principal components, distributional regression frameworks such as the Generalized Additive Models for Location, Scale and Shape (GAMLSS), and various types of deep learning approaches, including artificial neural networks (ANNs) and long short-term memory (LSTM) models. The atmospheric and climate models employed for hybrid forecasting can range from single models to large multi-model ensembles. For example, there are the North American Multi-Model Ensemble (NMME, Kirtman et al., 2014) and the Copernicus Climate Change Service (C3S) seasonal forecasting systems over sub-seasonal to seasonal timescales, or the Coupled Model Intercomparison Project (e.g. CMIP5-6) over decadal timescales. The dynamical predictors may include various model outputs such as meteorological forecasts with lead times of up to 14 days; initialized climate predictions with sub-seasonal to decadal lead times; sub-seasonal runoff predictions; and/or land surface or ocean state fields from the reanalyses used to initialize the climate system. Predictors are selected based on their ability to enhance hybrid forecast skill, such as traditional hydroclimate variables (e.g. precipitation, temperature, evapotranspiration) but also large-scale climate indices and teleconnections (e.g. DelSole and Shukla, 2009). Hybrid hydroclimatic forecasts and predictions have numerous operational and strategic applications, including water resources planning, reservoir inflow management (Tian et al., 2021; Essenfelder et al., 2020), surface water flooding (Rözer et al., 2021), flood risk mitigation, navigation (Meißner et al., 2017), and agricultural crop forecasting (Cao et al., 2022; Slater et al., 2022).

This paper provides an overview of recent developments and ongoing challenges in hybrid hydroclimatic forecasting. We seek to highlight the benefits of employing hybrid methods alongside or within traditional forecasting systems. Accordingly,

Table 2. Examples of hybrid forecasts of different hydroclimate variables and model types. Each example includes both a data-driven model and a dynamical weather or climate model. Examples are sorted by increasing time horizon. Hybrid model types are defined in Table 1 and acronyms are defined in Table 3.

Predictand	Data-driven model	Dynamical model	Hybrid type	Time horizon	Citation
River stage and inundation	LSTM	ECMWF HRES	Stat-dyn	1-6 days	Nevo et al. (2022)
Daily streamflow	BNN, SVR, GP, MLR	NOAA GFS	Stat-dyn	1-7 days	Rasouli et al. (2012)
Precipitation	RF	FV3GFS	Coupled	1-10 days	Watt-Meyer et al. (2021)
Precipitation extremes and flooding	Probability of exceeding thresholds	UKMO GloSea5, ECMWF	Serial	15 days	Richardson et al. (2020)
Biweekly temperature and precipitation	PLSR	CFSv2	Serial	2–3 & 3–4 weeks	Baker et al. (2020)
Seasonal streamflow	PCR & CCA	CFSv2 & ECHAM4.5	Stat-dyn	1 month	Sahu et al. (2017)
Monthly reservoir inflow	RF, GBM, ELM, M5-cubist, elastic net	FLOR	Stat-dyn	1 month	Tian et al. (2021)
Drought: seasonal SPI	Dynamic-LSTM	ECMWF SEAS5	Stat-dyn	3 months	Wu et al. (2022)
Seasonal tropical storm frequency	MLR	UKMO Glosea5	Stat-dyn	3 months	Kang and Elsner (2020)
Seasonal rainfall	ANN, MLR	UKMO GloSea5, ECMWF SEAS5	Stat-dyn	1-4 months	Golian et al. (2022)
Drought	Bayesian model based on copula functions	NMME (8 models)	Coupled	3-5 months	Madadgar et al. (2016)
Accumulated seasonal reservoir inflow	SVR, GP, LSTM, NLANN, DL	CMCC	Serial + stat-dyn	1-6 months	Essenfelder et al. (2020)
River discharge and surface water levels	MLR, LR, DT, RF, LSTM	ECMWF SEAS5; EFAS hydrological forecasts	Stat-dyn	1-7 months	Hauswirth et al. (2022)
Hurricane frequency and intensity	GAMLSS	NMME (6 models)	Stat-dyn	1-9 months	Villarini et al. (2019)
Seasonal runoff	PCR	NMME (7 models); ECWMF SEAS4	Stat-dyn	4-9 months	Lehner et al. (2017)
Hurricane frequency	Statistical emulator of dynamical atmospheric model	GFDL–CM2.1; NCEP–CFS	Stat-dyn	1-10 months	Vecchi et al. (2011)
Seasonal streamflow	GAMLSS	NMME (8 models)	Stat-dyn	1-10 months	Slater and Villarini (2018)
Monthly streamflow	FoGSS, CBaM	POAMA-M2.4	Serial	1-11 months	Bennett et al. (2016)
Seasonal flood magnitude	GAMLSS	5/8 CMIP5/6 GCMs	Stat-dyn.	2-5 years	Moulds et al. (2023)
Seasonal flood counts	Poisson regression	9/14 CMIP5 GCMs	Stat-dyn	1-10 years	Neri et al. (2019)
Daily streamflow	TCNN (& others)	4 GCMs from LOCA (CMIP5)	Serial + stat-dyn	Decades	Duan et al. (2020)
Flood magnitude	LSTM (+5 GHMs)	5 GCMs from ISIMIP-FT (CMIP5-6)	Serial	Decades	Liu et al. (2021)
Daily streamflow	DNN-PCE	10 GCMs (CMIP5)	Serial	Decades	Zhang et al. (2022)

in Section 2, we provide several in-depth examples of different approaches to hybrid hydroclimatic forecasting. In Section 3,

Table 3. Modelling acronyms referred to in the manuscript. Top box includes data-driven models & approaches; bottom box includes other acronyms used.

Acronym	Full name
ANN	Artificial neural network
BAMLSS	Bayesian additive models for location, scale and shape
BMA	Bayesian model averaging
BNN	Bayesian neural network
CBaM	Calibration, bridging and merging
CCA	Canonical correlation analysis
DL	Deep learning
DLNN	Deep-learning neural network
DNN-PCE	Deep neural network-based polynomial chaos expansion
DT	Decision tree
ELM	Extreme learning machine
FoGSS	Forecast guided stochastic scenarios
GAMLSS	Generalised additive models for location, scale and shape
GAN	Generative Adversarial Network
GBM	Gradient boosting machine
GP	Gaussian process
LR	Lasso regression
LSTM	Long short-term memory
ML	Machine learning
MLR	Multiple linear regression
NLANN	Non-linear autoregressive neural network
PCR	Principal component regression
PLSR	Partial least squares regression
RF	Random forest
SVM	Support vector machine
SVR	Support vector regression
TCNN	Temporal convolutional neural network
CMIP5&6	Coupled model intercomparison project phases 5 and 6
FV3GFS	Finite-Volume Cubed-Sphere Global Forecast System (global atmospheric model)
GCM	Global climate model
GHM	Global hydrological model
ISIMIP	Inter-sectoral impact model intercomparison project
PREVAH	PREcipitation-Runoff-EVApo-transpiration HRU Model)
RCP8.5	Representative Concentration Pathway 8.5 (high-emissions warming scenario)

we discuss the key strengths of hybrid models, followed by ongoing challenges and future research opportunities in Section 4. We close with some concluding remarks in Section 5.

2 Hybrid forecasting examples

Here we provide examples of the statistical-dynamical, serial, and coupled approaches outlined in Figure 1 and Table 1.

2.1 Statistical-dynamical hybrid forecasts

In the case of short-term hybrid forecasts, which focus on outlook horizons of hours to weeks driven by dynamical meteorological models, hybrid approaches offer potential for addressing the challenge of forecasting extreme events, such as floods from convective rainfall (Speight et al., 2021). In these situations, the time taken to transfer data between meteorological and hydrological organisations and the run time of physics-based models can be restrictive. In contrast, the strengths of ML are the small number of input parameters making the models easy to develop, quick to run, and accurate for short lead-time events (Piadeh et al., 2022). In regions where access to hydrological and inundation forecasts is limited, data-driven models offer promising alternatives for flood forecasting (e.g. Nevo et al., 2022) and show potential to overcome limitations of data scarcity (Kratzert et al., 2019a; Feng et al., 2021). At 1-7 day lead times, Rasouli et al. (2012) found that ML models outperform MLR (Tables 2-3). At the shortest lead times, their hybrid approach worked best when it was driven by observations and the NOAA Global Forecasting System (GFS) model output, and at longer lead times when driven by a combination of local observations and climate indices. The potential of ML as a means to post-process dynamical forecasts and produce warning scenarios for convective weather is also emerging (e.g. Moon et al., 2019; Flora et al., 2021) but has not yet been widely utilised as input to hydrological models. For hydrologic forecasts, ML is highly successful in assimilating recent observations of streamflow to improve near-term daily forecasts of streamflow (Feng et al., 2020) and soil moisture (Fang and Shen, 2020b). In some cases, machine learning can ingest near-real-time data without the need for backwards methods like data assimilation, since any data stream can be fed directly into the model as inputs, as long as at least some samples from each input data stream are available during training. It is also possible to perform more traditional types of data assimilation on or with ML models – for example variational assimilation can be done by leveraging the same partial gradients in the models that are required for backpropagation (Nearing et al., 2022).

At the sub-seasonal to decadal timescale, climate model predictions are often used to drive statistical or ML models. A simple example of a hybrid statistical-dynamical model is one that employs the predictions of precipitation or temperature from a climate model as predictors within a regression model, where the predictand can be a hydroclimatic variable such as streamflow magnitude (e.g. Slater et al., 2019) or flood duration (Neri et al., 2020). Schlef et al. (2021) describe this approach as an 'informed-parameter approach' in which the parameters of the flood distribution can be conditioned on time-varying covariates such as time, climate indices, infrastructure development indices, or land use indices. For example, distributional regression models can be used to predict seasonal discharge. To illustrate the approach, we consider a 9000 km^2 catchment that has experienced rapid expansion of the agricultural land area over the 20^{th} century (Figure 2). Two lumped covariates are employed to predict the seasonal maximum of mean daily streamflow in each year: the basin-averaged total seasonal precipitation and the harvested corn and soybean acreage in the same season. The model employs a two-parameter gamma distribution, and the entire streamflow distribution is computed for each timestep. The model is trained over the historical period using climate observations or forecasts, model parameters are extracted, and the streamflow forecast is based on those parameters and the dynamical predictions of the covariates obtained from an ensemble of climate models. Once new observations become available, the model can be retrained, updating the model parameters. A different model can be developed for each season,

initialization time (e.g. 0.5, 5.5 and 9.5 months ahead of a given season), and quantile of the predicted discharge distribution. This example shows how a simple statistical model can be used to produce sub-seasonal to seasonal streamflow forecasts. The skill of such a scheme might be improved by post-processing the ensemble of climate predictions used to drive the model.

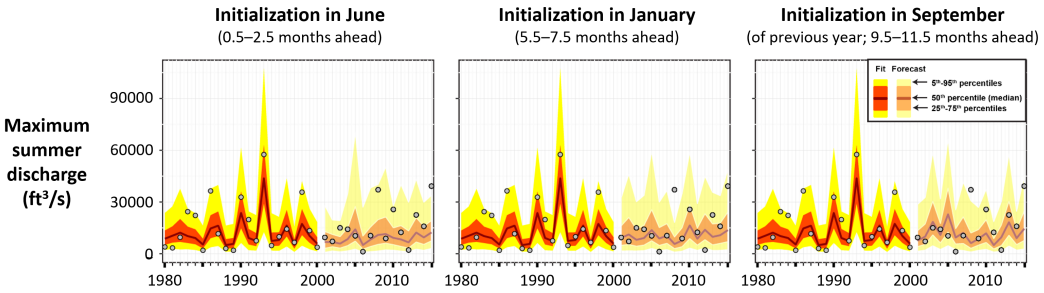

Figure 2. Example of seasonal hybrid forecasting system for maximum summer discharge at one stream gauge, using seasonal climate forecasts from 8 climate models (94 members) of the NMME to drive a distributional regression model of streamflow. The maximum summer discharge is the largest of the 92 daily values in the summer (JJA) period. The time series shows the model fit (1980-2000) and forecast (2001-2015) against the observations of maximum summer daily streamflow (grey circles). Initialization times are 0.5, 5.5 and 9.5 months ahead of the summer season. For example, 'initialization in June' uses climate forecasts with 0.5-month lead for June, 1.5-month lead for July, and 2.5-month lead for August to compute the summer streamflow, while 'initialization in September' includes forecasts initialized 9.5, 10.5 and 11.5 months ahead in the previous year. Modified from Slater et al. (2019).

Seasonal forecasts of diverse hydroclimatic variables such as precipitation, evaporation, sea water level, sea level pressure or large-scale climate indices have also been used to drive hybrid models to predict variables such as precipitation (Madadgar et al., 2016) and tropical cyclone activity (Sabeerali et al., 2022; Murakami et al., 2016). For instance, atmosphere-ocean teleconnections obtained from the NMME – including the Pacific Decadal Oscillation (PDO), Multivariate ENSO Index (MEI), and Atlantic Multidecadal Oscillation (AMO) – were used to successfully predict seasonal precipitation anomalies in the southwestern USA using a statistical Bayesian-based model (Madadgar et al., 2016). Hybrid methods can also be trained on large model ensembles to capture non-linear interactions between predictor variables. For instance, Gibson et al. (2021) trained ML models for seasonal precipitation forecasts in the western USA on a large historical climate model ensemble of atmospheric and oceanic conditions (i.e. on thousands of seasons of simulations from the Community Earth System Model Large Ensemble, CESM-LENS). The same trained models were then tested by using observational data over 1980-2020. The resulting ML-based approach performed as well as, if not better than, seasonal NMME forecasts, and the physical processes could be interpreted using ML interpretability plots, highlighting the most important variables influencing a given forecast. For Ireland, Golian et al. (2022) found that MLR and ANN models applied to hindcasts of mean sea level pressure from GloSea5 and SEAS5 produced skillful forecasts of winter [DJF] and summer [JJA] precipitation for lead times of up to four months, with the ANN outperforming MLR for both seasons and all lead times. A study over the Netherlands using streamflow, precipitation, and evaporation found that the hybrid ML approach outperformed climatological reference forecasts by approximately 60%

and 80% for streamflow and surface water level, respectively, using various machine learning models (Hauswirth et al., 2022). Another study showed that predictions of large-scale indices by the CFSv2 model could be used to successfully predict the frequency of tropical cyclones in the Bay of Bengal using principal component regression (Sabeerali et al., 2022).

Statistical-dynamical approaches can also be deployed for longer horizons such as decadal streamflow predictions (e.g. Neri et al., 2019), and data-driven techniques are proving successful for enhancing the skill of the decadal climate predictions, with consequent benefits for climate-linked variables such as streamflow. Decadal forecast skill can be increased by 'mode-matching', which consists of sub-selecting the individual members from a large climate model ensemble of decadal predictions that best represent the multiyear temporal variability of a relevant large-scale mode of climate variability (Smith et al., 2020; Moulds et al., 2023). Large climate ensembles can be pre-processed to select members which are skilful at a given time, and the improved predictions can then be supplied to a statistical modelling framework to predict seasonal streamflow quantiles (Moulds et al., 2023).

2.2 Serial hybrid forecasts

2.2.1 Serial pre- and post-processing of hydroclimate predictions using data-driven approaches

Hybrid approaches often include pre-/post-processing of inputs and outputs at different stages of the predictive model. Pre-processing refers to techniques for enhancing the signal and removing systematic biases of the data inputs, such as the dynamical climate simulations, while post-processing refers to techniques for refining and correcting model outputs. Depending on the point of reference, the same technique can be considered as either pre-/post-processing. It is important to point out that pre-/post-processing is also used as a routine add-on to traditional forecasting systems (e.g. driving a hydrological model with pre-processed climate predictions) and here we focus on approaches that go beyond the traditional setup. The strength of hybrid approaches lies in their ability to incorporate such corrections directly within hybrid modelling frameworks.

Hybrid models often include a data-driven component which downscales low-resolution climate model simulations to reduce bias and make the outputs more skillful at the local scale. For instance, Generative Adversarial Networks (GANs) have been used to spatially downscale precipitation forecasts (Harris et al., 2022; Pan et al., 2022) to capture complex joint distributions between precipitation and initial climate conditions from climate simulations. At the decadal timescale, linear and kernel regression can be used to enhance climate predictions (Salvi et al., 2017a, b). Random Forest (RF) models can be trained to map low-resolution climate model predictions to high resolution values (Anderson and Lucas, 2018). Regardless of the algorithm used, once the mapping from low-resolution to high-resolution values has been learned, data-driven models can be applied to a much larger number of model simulations to produce an ensemble of high-resolution outputs at a much lower computational cost than running a dynamical model at an equivalent resolution. Another example is the use of data-driven methods to reduce the degrees of freedom in data, for instance through discrete or empirical wavelet transforms (Mosavi et al., 2018).

Data-driven approaches can also be applied directly to post-process the hydrological forecasts. Bennett et al. (2021a) deployed an ERRIS (error reduction and representation in stages) error model to directly correct errors in streamflow prediction

up to 168 hours ahead (i.e. maximum lead time of 7 days). Such approaches can be especially beneficial for longer forecast horizons. For instance, a Gaussian Process (GP) model was trained to post-process weekly tercile forecasts of runoff and soil moisture from a Swiss conceptual hydrological model PREVAH, and showed improvements in the forecast skill up to 4 weeks ahead (Bogner et al., 2022). McInerney et al. (2022) developed a daily Multi-Temporal Hydrological Residual Error (MuTHRE) statistical model to seamlessly transform daily streamflow forecasts to time scales ranging from daily, weekly, fortnightly to monthly. This one-model-for-all-scales approach is a novel take on the potential of the hybrid forecasting system. LSTMs can also be used to post-process outputs from physics-based models, such as long-term streamflow projections (Liu et al., 2021) and streamflow simulations (Frame et al., 2021) to make them more realistic. Liu et al. (2021) implemented a physics-informed approach to post-process the streamflow projections from GCMs, GHMs and the Catchment-based Macro-scale Floodplain model (CaMa-Flood). The LSTMs were trained to learn a relationship between simulated streamflow (from the physics-based model GHMs-CaMa-Flood), basin averaged daily precipitation, temperature, windspeed and observed streamflow. The LSTM model can thus be perceived as a post-processor which aims to constrain (i.e. reduce the uncertainty of) the streamflow simulations from the physics-based model. This post-processing approach improved the simulations for the reference period, and was then successfully applied to project streamflow over the future period. However, the authors concede that this LSTM-based post-processor is still subject to the same limitations as other post-processing methods, such as the assumption of stationarity in the parameters of the post-processing method. Frame et al. (2021) similarly employed an LSTM to post-process the outputs from the physics-based US National Water Model (NWM). They implemented two variants of the post-processing method, alongside an LSTM forced with atmospheric inputs only (i.e. without any NWM inputs). The authors showed that the routing scheme and the land surface component of the NWM introduced timing and mass balance errors in the simulations. Thus, in some cases, it would be preferable to simply use an LSTM model that can simulate streamflow from atmospheric forcings only (without any NWM inputs), to avoid propagating errors from the NWM to the streamflow prediction.

Data-driven models can enhance the signal of predictors by generating an ensemble (by pooling) of different climate model predictions (Troin et al., 2021). A common approach to incorporate an ensemble of climate model predictions (within a statistical, ML, or hydrological model) is to assume that predictions from each ensemble member are equally likely. However, owing to varying model skill, as well as a lack of independence amongst some models, the assumption of equal likelihood can be compromised. Hence, hybrid forecasting can be used to combine ensembles in more intelligent ways by accounting for the varying information content of ensemble members. Statistical ensembling/post-processing of climate model ensemble outputs can improve forecast skill at relatively low computational cost. For instance, Grönquist et al. (2021) applied a deep neural network to ensemble predictions to improve forecast skill and reduce the computational requirements of the forecast system. Massoud et al. (2020) applied Bayesian Model Averaging (BMA) to weight models according to their skill at reproducing observations. They show the weighted ensemble average skill for the contiguous Unites States exceeds that of the conventional ensemble average, with better constrained uncertainty estimates. Bayesian updating can also be applied to enhance the skill of a multi-model ensemble of GCMs such as the NMME for different seasons or lead times (e.g. Slater et al., 2017). Bayesian updating provides the best results when the raw GCM predictions have high skill to begin with, such as SST-based ENSO forecasts (Zhang et al., 2017). Post-processing hydrological forecasts (instead of climate forecasts) is another application of BMA.

Hemri et al. (2013) demonstrated how such an approach can be deployed to improve the skill of a conceptual runoff forecast by pooling four separate runoff forecasts forced with different lead times (24-hr, 72-hr, 120-hr, and 240-hr) and ensemble members (1, 1, 16, and 51, respectively) in a Swiss catchment.

2.2.2 Serial hybrid forecasts that include a hydrological model

Hybrid forecasting systems that include a conceptual hydrological model try to combine the strengths of data-driven and conceptual models, driven with dynamical predictions. For instance, Humphrey et al. (2016) used a combination of historical observations and downscaled dynamical forecasts of rainfall and PET in southern Australia from POAMA to drive the conceptual rainfall-runoff model GR4J (Perrin et al., 2003). The simulated soil moisture from GR4J was separately used to drive a Bayesian ANN model to predict streamflow (hybrid approach). They showed that the hybrid model performed better than either the GR4J model or the Bayesian neural network alone. A number of studies have coupled conceptual models and data-driven models, but without necessarily integrating dynamical weather or climate predictions (this would be the next step in developing a hybrid forecasting system). Both Anctil et al. (2004) and Kumanlioglu and Fistikoglu (2019) replaced the routing component of the GR4J model with an ANN to predict streamflow in catchments in France, the USA and Turkey. These studies concluded that the hybrid model was superior to a purely ML model. Other conceptual hydrological models have also been used in hybrid frameworks. For example, Mohammadi et al. (2021) used two conceptual models, HBV (Bergström, 1976) and NRECA (Crawford and Thurin, 1981) to provide inputs to support vector machines (SVM) and adaptive neuro-fuzzy inference system (ANFIS), to build seven variants of hybrid models. They tested and compared the hybrids as well as the individual models (HBV, NRECA, SVM and ANFIS) on four sub-basins of the Pemali Comal River Basin, Indonesia, and again found the hybrid models performed best in terms of RMSE, R^2 and MAE. Other studies on hybrid modeling using the HBV model include Nilsson et al. (2006) and Ren et al. (2018). They both used different variables computed by HBV (e.g. soil moisture, snowmelt) as inputs to ANNs. Okkan et al. (2021) outline that in most hybrid modeling frameworks, variables computed by the conceptual model are used as inputs to a data-driven model, which necessarily increases computation time. They also note that although there could potentially be interactions between the parameters of the conceptual models and those of the data-driven model, those interactions often go unaccounted for because the two models are calibrated separately. In the context of monthly rainfall-runoff modelling, they proposed to address these two common shortcomings of hybrid models by coupling the two models and performing their calibration jointly.

2.3 Coupled or parallel hybrid models

In the case of coupled hybrid models, a data-driven model and a physics-based model can be run in parallel, sometimes replacing a component of the dynamical model with a data-driven model or combining different types of model predictions. Madadgar et al. (2016) combined the seasonal precipitation predictions from an ensemble of dynamical models (99 members from the NMME) with the precipitation predictions from a statistical model (using copulas to describe the relationship between three large-scale climate indices and precipitation). They used an Expert Advice algorithm to link the dynamical and statistical predictions to obtain improved precipitation predictions over the southwestern USA, as illustrated in Figure 3.

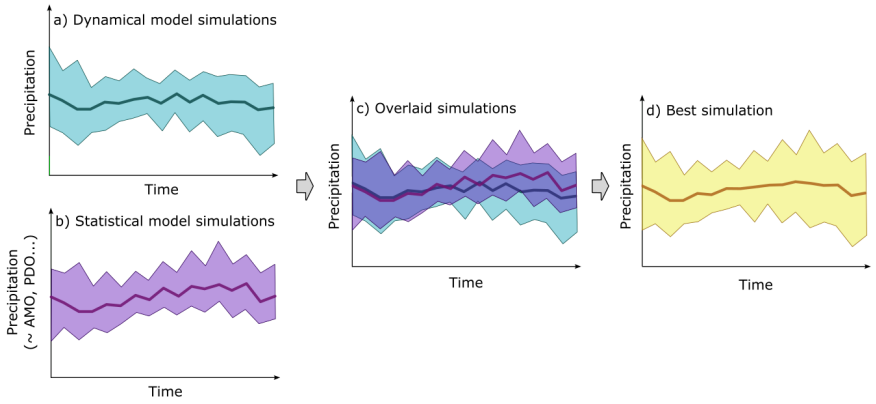

Figure 3. Example of a coupled hybrid system for predicting seasonal precipitation several months ahead. (a) Ensemble of precipitation predictions from a dynamical multi-model ensemble such as the NMME. Ribbon indicates the full distribution of model members; dark line indicates the mean prediction. (b) Ensemble of statistical precipitation predictions. (c) Both ensembles are overlaid. (d) The two ensembles are blended using a data-driven approach, such as an Expert Advice algorithm, which assigns weights to the different ensemble members based on their performance during training and computes the weighted average prediction. The resulting ensemble mean (orange line) outperforms that of the separate dynamical and statistical predictions. Adapted from Madadgar et al. (2016).

Coupled hybrid models can also employ a data-driven model to combine other types of dynamical predictions in parallel, such as dynamical meteorological and hydrological predictions. In southern Switzerland, five ML models were trained to predict monthly total hydropower production by combining precipitation, temperature, radiation, and windspeed forecasts from a dynamical meteorological model with runoff from a conceptual hydrological model (Bogner et al., 2019). Day of the week and holiday information were provided to the ML models as additional information to further enhance the prediction.

A third example of a coupled hybrid approach is when data-driven models are employed during the dynamical climate model simulations to correct model biases (e.g. Watt-Meyer et al., 2021). A RF model coupled to an atmospheric model (FV3GFS) can correct temperature, specific humidity and horizontal winds at each timestep, bringing the coupled model in line with observations. This was shown to reduce annual-mean precipitation biases by around 20%, with particular improvements in the simulation of rainfall over high mountains (Watt-Meyer et al., 2021). A similar approach was used by Bretherton et al. (2022) to nudge the output of a low-resolution climate model towards the coarsened output of a high-resolution climate model.

3 Strengths of hybrid forecasting

Hybrid methods offer various strengths, as summarized in Figure 4. These include benefits related to the higher performance of ML models (in terms of bias and error minimisation), the ability to easily blend outputs from climate multi-model ensembles, integrating large datasets, combining multiple sources of predictability to enhance predictive skill, improved speed and operational convenience. These strengths are discussed in more detail below.

3.1 ML model performance and bias minimization

Recent work has demonstrated the ability of ML models to outperform traditional hydrological models (e.g. Fang et al., 2017; Kratzert et al., 2019b; Feng et al., 2020; Fang and Shen, 2020a; Lees et al., 2021). In one of the most comprehensive studies to date, Mai et al. (2022) compared 13 locally- and globally-calibrated models (including ML, lumped and gridded models) in terms of their ability to simulate streamflow, actual evapotranspiration, surface soil moisture and snow water equivalent in the Great Lakes region. They found that the ML model outperformed the traditional hydrological models in all experiments. This finding extends to ungauged catchments: Kratzert et al. (2019a) found an out-of-sample LSTM performed better than the calibrated SAC-SMA (the conceptual model used by the US River Forecast Centers) and the U.S. National Water Model, which is less calibrated. Golian et al. (2021) found that random forests worked best at regionalizing the parameters of the GR6J conceptual model for low flow prediction in ungauged Irish catchments. Such work has shown the potential of hybrid methods to address the longstanding hydrological challenge of prediction in ungauged basins (e.g. Sivapalan, 2003). The next step is to move from simulation to prediction.

Hybrid models combining ML and climate predictions also tend to outperform the raw dynamical forecasts from climate models. Wu et al. (2022), for instance, developed a hybrid drought-forecasting model of the 3-month Standardised Precipitation Index (SPI). They used random forest models to post-process ECMWF SEAS5 predictions of geopotential height, sea level pressure and air temperature, and supplied the output to an LSTM model to predict the 3-month SPI. They found that the SPI predictions from these hybrid models outperformed the predictions of SPI obtained from the raw model outputs. For prediction purposes, hybrid models have the advantage of being able to minimize biases that exist within GCM outputs or that might be otherwise introduced within a hydrological modelling chain. By training a hybrid model directly on the climate model forecasts/predictions, rather than on observations, the biases are automatically accounted for within the model (e.g. Slater and Villarini, 2018). This approach is similar to that of model output statistics (MOS) long used by the weather forecasting community (Glahn and Lowry, 1972) and in seasonal hydrological predictions (Schick et al., 2018). For instance, if a climate model tends to overpredict winter rainfall, this bias is accounted for directly in the streamflow predictions, given that the model is trained using the same winter rainfall forecasts (assuming a constant bias).

Hybrid models may benefit from a wide range of statistical advances for enhancing the skill of hydroclimate predictions. Since a hybrid system is based on a data-driven model, it is straightforward to incorporate statistical 'upgrades', such as ensembling the outputs of multiple climate or Earth System Models (Duan et al., 2019). One such example is the addition of an error model onto Ensemble Streamflow Prediction (ESP) forecasts to enable prediction in ephemeral rivers (Bennett et al., 2021b). In a hybrid system, one may easily integrate the predictions from multi-model ensembles with over 50 or 100 model members as covariates (Gibson et al., 2021; Slater and Villarini, 2018). Increasing the number and diversity of climate models included within a hydrological predictive model enhances confidence in the hydrological model spread. By blending multi-model ensembles intelligently one can further reduce uncertainty. In a hybrid system, for instance, one can incorporate time-varying weights for the dynamical predictions, such as Bayesian updating - varying model weight per month and lead time (Slater et al., 2017). ML models especially can learn space-time variable input weighting directly (Kratzert et al., 2021).

Similarly, many post-processing methods can be applied to weather and climate inputs or the hydrological outputs to enhance skill (Monhart et al., 2019; Bogner et al., 2022).

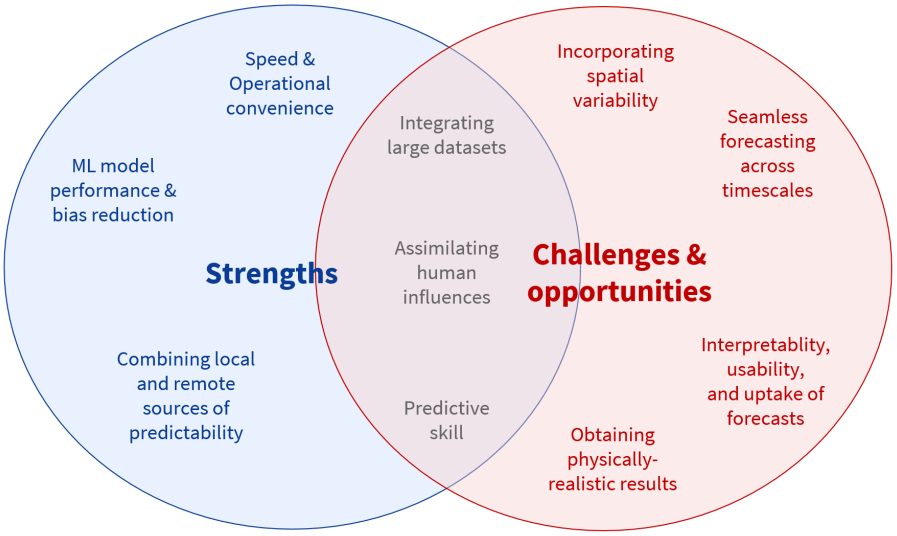

Figure 4. Strengths, challenges and opportunities of hybrid hydroclimate prediction systems, as discussed in Sections 3 and 4.

3.2 Combining local and remote sources of predictability with varying time-horizons

One under-researched but promising aspect of hybrid models is their ability to combine different sources of predictability over a continuum of time horizons. Hybrid models can easily make use of different predictors chosen on a sound physical basis (such as climate indices, precipitation, air pressure, snowfall) without explicitly describing the processes and equations. This makes it much easier to explore information from new sources and improve models, and has the potential to widen information access to climate-affected populations. Including additional inputs can also produce marked improvements in model quality. Chang et al. (2022, under review) used seven weather regime indices (based on the 500 hPa geopotential height) with a Gaussian Process ML model to post-process sub-seasonal hydrological forecasts, alongside runoff, soil moisture, baseflow, and snowmelt in Switzerland. The results showed that the additional input of weather regime indices improved the forecast skill especially in the mountainous catchments and over longer lead times, where skill was difficult to improve without any additional information. The conceptual hydrological model would not have been able to take weather regime indices as input, but by including them in the post-processing ML model as part of the hybrid setup, it was possible to explore the connection between large scale weather regimes and local hydrological conditions to improve the forecast skill.

As multiple predictor variables can be included within a statistical or ML model, it is feasible to combine predictors that have very different time-varying impacts, such as reservoir management decisions or initial hydrological conditions impacting the short term, versus annual-to-multidecadal climate oscillations for longer-term predictability. For instance, Tian et al. (2021) present a reservoir inflow forecasting framework combining a suite of different ML models (including gradient boosting ma-

chine, random forests, and elastic net) with climate model outputs from the FLOR model, for reservoirs in the Upper Colorado River Basin. They also included soil moisture and evaporation to represent antecedent conditions, which significantly improved the forecasts of reservoir inflow. Ouyang et al. (2021) used a dataset of >3000 basins across the USA and found that basins with small and medium reservoirs behaved differently from the reference basins but could be well simulated by a LSTM model with input attributes describing basin-lumped reservoir statistics.

Large-scale climate indices or modes can also be combined with other predictors. For instance, Madadgar et al. (2016) predicted seasonal precipitation using large-scale climate indices: the PDO, the MEI, and the AMO, computed from outputs of the 99 ensemble members of the NMME. The approach enhanced the skill of the seasonal forecasts by 5-60% in comparison with the raw NMME precipitation forecasts, especially for negative rainfall anomalies. Similarly, Rasouli et al. (2012) forecasted daily streamflow in a river catchment 1-7 days ahead by employing weather forecasts from the NOAA GFS model within a variety of machine learning models. They combined observations with the model outputs and a range of large-scale climate indices representing ENSO, the Pacific-North American teleconnection (PNA), the Arctic Oscillation (AO) and the North Atlantic Oscillation (NAO). Lastly, Li et al. (2022) used forecasts of the intraseasonal oscillation (ISO), an important mode of sub-seasonal predictability for seasonal rainfall, to force a Bayesian hierarchical model predicting sub-seasonal precipitation during the boreal summer monsoon season in different regions of China.

Given the diversity of potential inputs to hybrid forecasting systems, exploratory data analysis to identify correlations between hydrologic variables and climate patterns over different time horizons is an important step during model development. Hagen et al. (2021) employed ML to identify the most relevant large-scale climate indices for daily streamflow forecasting. They provided an overview of studies that have employed large-scale climate indices and climate variables (such as sea level pressure, sea surface temperature, specific and relative humidity) within ML models for daily, monthly and seasonal streamflow modelling. Beyond the use of pre-defined climate indices, it is possible to identify tailored, site-specific climate indices from big data and incorporate them in the modelling chain. For instance, Renard and Thyer (2019) described a method that avoids relying on standard climate indices and instead suggests that the most relevant climate indices in a given location are effectively unknown (they are 'hidden') and can be estimated directly from observations. The authors used a Bayesian hierarchical model for flood occurrence, with hidden climate indices treated as latent variables. They identified the hidden climate indices and then showed their correlation with atmospheric climate variables (geopotential height, zonal westerly wind, but also more distant teleconnections using convective available potential energy and meridional wind). These indices explain the occurrence of flood-rich and flood-poor periods in the historical record. Such an approach could be employed using climate model outputs to develop skillful hybrid forecasts.

Related to the different time-horizons of the predictors is also the ability to design hybrid forecasting systems which dynamically update when new information (e.g. observations or climate hindcasts) become available. For instance, a statistical model can be updated iteratively over time to track the evolution of nonstationary predictor-predictand relationships. Such approaches incorporate new observations as they become available and update the model parameters (e.g. Slater et al., 2019). Nearing et al. (2022) developed a data assimilation approach for LSTM models that leverages tensor network gradients to assimilate real-time observation data. To date, very little has been published using such methods.

3.3 Integrating large datasets

One perceived challenge of hybrid approaches is the requirement for large amounts of training data to constrain models compared with physics-based or conceptual models. Previously, it was felt that the information requirement of data-driven approaches might hinder their applicability in catchments with limited data (e.g. ungauged basins). Although this might have been true in the past, the increasing availability of large-scale hydroclimatic datasets such as remote sensing data is turning this potential challenge into a new opportunity. A data-driven model can be trained on the same data as a conceptual model, and will tend to out-perform physics-based models, on average (and even more so with large datasets; see Fang et al., 2022). This advantage is partly due to the fact that data-driven models are unconstrained by mass and energy balance rules that force process-models to compensate for erroneous inputs, which data-driven models can instead optimize against. Data-driven models 'learn' process relationships and model structures rather than enforce prescribed ones, which may make them more flexible and generalizable. Large training datasets tend to be useful for ML but less so for physics-based models, for these reasons. The ability to leverage large datasets effectively is a strength of ML, and in particular for ungauged basins, where several studies have shown that ML models tend to have higher accuracy, on average, than physics-based models calibrated in gauged basins (e.g. Kratzert et al., 2019a). There is, in fact, a 'data synergy' effect, where data of greater diversity lead to better models, according to a systematic study of LSTM models for either streamflow or soil moisture (Fang et al., 2022). With conceptual and process-based models, accuracy can be lost when performing regional (as opposed to basin-specific) calibration, and the lack of calibration data typically results in poor-quality predictions (training on longer periods leads to superior results – see Bogner et al. (2022)). In contrast, with hybrid models, strong performance can be achieved when training the models on global datasets, and accuracy is gained when performing regional calibration.

Since long (50-year +) hydroclimatic time series data are not available everywhere (Krabbenhoft, 2022), methods are required that draw on pooled multi-site approaches with similar catchment and climate characteristics (Kratzert et al., 2019a). For instance, Nearing et al. (2021) show a comparison using pooled vs unpooled data for streamflow estimation and found the former was better, even for gauged catchments, and allowed for prediction in ungauged catchments. There are, however, few studies combining LSTM methods with climate model forecasts for long-term (sub-seasonal to decadal) prediction, especially in ungauged catchments. Such models may start to emerge with the growing availability of observational training datasets, such as the national 'CAMELS' datasets (available for the USA, United Kingdom, Chile, Brazil, Australia, France, and soon Switzerland, e.g. Newman et al., 2015; Addor et al., 2017; Coxon et al., 2020) and international 'Caravan' streamflow dataset (Kratzert et al., 2023). However, real-time data are currently still difficult to access for developing predictive models.

One way to circumvent the lack of observational training data and the low predictability of GCMs is by integrating a range of other types of predictors in hybrid models. This may include sources of remotely sensed measurements such as snow, soil moisture, land cover, surface water extent, water storage or evapotranspiration to provide better information about initial states (e.g. Jörg-Hess et al., 2015). There are many different global datasets now available that can be drawn on using cloud-based geospatial analysis platforms such as Google Earth Engine, as was the case for the creation of an open-source community streamflow dataset (Kratzert et al., 2023). Overall, the forecasting landscape is becoming increasingly complex,

with a growing number of forecasting systems and datasets potentially overwhelming users. Hybrid forecasting could help to address this challenge, with hybrid workflows providing a set of tools and data that forecasters could mix and match to address their own forecasting needs.

3.4 Speed and operational convenience

A key advantage of statistical or hybrid methods is their speed and computational efficiency. For instance, the calibration of the GloFAS system with an Evolutionary Algorithm (EA) in 2018 required approximately 6 hours to calibrate each one of 1000s of streamflow stations on a 12-core PC, depending on the number of generations needed before the improvement criterion was met (Hirpa et al., 2018). Training deep learning (DL) models is now orders of magnitude cheaper. For example, it took about 10 hours in 2021 to train an ensemble of Long Short-Term Memory (LSTM) networks on a single NVIDIA V100 GPU using two decades of daily data from 518 basins in the CAMELS-GB dataset (Lees et al., 2021), i.e. about 70 seconds per basin. This means that training a high-quality DL model for hundreds of basins is feasible using a standard workstation (or even a GPU-enabled laptop with sufficient memory), while calibrating a conceptual or process-based model over hundreds of basins requires either months of runtime or an HPC facility. The training time depends on the computing power, number of locations and amount of data involved, compiler, and optimization. While deep learning methods such as LSTMs can take several hours to train (e.g. Lees et al., 2021), they have the significant advantage that one model is trained on multiple sites (although the fitted model can then be fine-tuned to a specific site). A differentiable ML-based parameter learning scheme can be trained on satellite-based soil moisture observations for the entire continental USA with one GPU in under one hour, but the conventional approach would take a cluster machine of 100 CPUs 2-to-3 days to calibrate the model (Tsai et al., 2021).

This efficiency has advantages for water managers. In a traditional setting with limited computational resources, water managers need to quickly run different scenarios (Scher et al., 2021). For instance, the UK Flood Forecasting Centre will produce a 'reasonable worst case' and a 'best estimate' based on the most likely scenario (see Met Office, Environment Agency and Flood Forecasting Centre (2013)) ahead of a flood event (Arnal et al., 2020). Using all available deterministic and ensemble forecast products alongside expert assessment from the chief forecaster they will decide what the reasonable worst case is likely to be. These outputs are used to inform the flood guidance statement and the Environment Agency then uses these scenarios to run their catchment models (Pilling et al., 2016). The speed of data-driven approaches in comparison with these more traditional physics-based modelling approaches could prove beneficial for users wishing to run multiple scenarios quickly. Hybrid methods may shorten the traditional forecasting approach by going 'end-to-end', potentially skipping out some of the intermediary steps in a conventional modelling chain, such as downscaling, bias correction and hydrological modelling. This offers significant potential for applications where the run time of physically based models limits the ability to provide forecasts with a useful lead time for action – such as forecasts of pluvial floods Rözer et al. (2021) or flash floods.

The efficiency of hybrid models may also be helpful in generating faster research cycles for model improvements (i.e. setting up an upgraded system and releasing hindcasts for testing) relative to traditional approaches. Model upgrades for dynamical systems usually take a very long time because the model has to be re-calibrated and a set of X (e.g. 30) years of hindcast data must be produced to quantify the impact of the changes to the system.

Lastly, hybrid systems can be used to develop customized climate services. For instance, Essenfelder et al. (2020) use data-driven methods to predict seasonal reservoir inflows for hydropower plants. The information is made easily accessible online to support decision-makers in hydropower production. Such approaches can be designed to be replicated globally as a climate service, provided there are suitable data for training, and by developing transferable rule sets. Bennett et al. (2016, p.8239) also highlight the importance of operational convenience and the advantages of combining 'the convenience of stochastic scenarios with the skill of a modern forecasting system'. Their method enhances precipitation forecasts necessary for streamflow forecasting through post-processing - by reducing the biases, correcting the reliability, and maximising the forecast signal.

4 Key challenges and opportunities of hybrid forecasting

Beside the strengths of hybrid methods, there are challenges and research priorities to be tackled. As hybrid forecasts and predictions rely on data-driven models, they inevitably inherit some of the limitations of these techniques. Frequently-cited limitations of ML models include the requirement for large datasets and issues associated with the 'curse of dimensionality', namely data sparsity (i.e. when there are too few data points relative to the number of dimensions), multicollinearity of the variables, multiple testing (leading to an increased number of false positives), and overfitting (Altman and Krzywinski, 2018). There is also the difficulty of obtaining physically plausible results for previously 'unseen' extremes that are larger than those seen in the observational record; however, new research suggests that ML models may provide results that are more physically plausible than physics-based and conceptual models when data are biased (Frame et al., 2022b). Further challenges for improving the skill of hybrid models include data assimilation, physics-guided ML designs, assimilation of human influences, model optimisation, ensembling, and hybridization, where models are merged with other methods (including simulations and physical models, e.g. Mosavi et al., 2018). While some of the difficulties associated with large sample sizes apply less for seasonal to decadal hybrid forecasting, where the sample sizes can be much smaller (often near 100 values) than the sample sizes for shorter ranges (thousands or more), the small sample sizes present a challenge for model training. Thus, a range of different challenges may apply depending on the forecasting horizon and data required.

4.1 Obtaining physically realistic results

One important challenge of hybrid models is the need to produce physically-plausible or explainable forecasts in unseen extreme conditions such as severe floods, droughts, intense heatwaves and tropical storms. This is particularly important as new weather records are being set in different parts of the world, and models must produce credible predictions under extreme forcing conditions. Although it has sometimes been suggested that data-driven models might be less suited to extrapolation to out-of-sample conditions than physics-based models due to the lack of physical understanding (e.g. Reichstein et al., 2019), recent work tackled the question of whether modern LSTMs could predict events larger than those seen in the training data for a particular catchment. The authors found that the LSTM could predict 'unseen' streamflow extremes, and did this better than the physics-based models that were used in the study (Frame et al., 2022a). It is now increasingly recognised that one of the advantages of data-driven models is their flexibility, allowing them to find unexpected patterns in the data. Thus, there are

emerging synergies between data-driven and physics-based approaches, since the former can enhance the performance of the latter, e.g. by learning the parameterizations required for the physical models from large datasets or analysing the patterns of error from the physical models (Reichstein et al., 2019).

One emerging route for hybrid models is to employ physics-guided or theory-guided ML designs that explicitly observe the law of conservation of mass. Such approaches seek to integrate physical knowledge within the data-driven models to take advantage of the strengths of both. For instance, Hoedt et al. (2021) created an LSTM architecture that obeys conservation laws, and these laws can also be used to guide physical interpretation of model outcomes. Although there have been considerable methodological advances in interpreting neural networks (e.g. Wilby et al., 2003; Toms et al., 2020; Lees et al., 2022), physics-guided ML approaches (also referred to as physics-informed, physics-aware, or theory-guided approaches) still require further development. As alluded to earlier, the presence of data errors in observed hydroclimate records means that an unconstrained ML performs better than a physics-guided ML model because of the ability to learn and account for data errors (Beven, 2020; Frame et al., 2022b), including heteroscedastic and nonstationary data errors (Kratzert et al., 2021).

Another new development is differentiable, learnable physics-based models that can approach the performance of ML models but also output internal physical variables such as evapotranspiration and soil moisture (Feng et al., 2022b; Shen et al., 2023). Tsai et al. (2021) first demonstrated the ability of connected neural networks to provide physical parameter sets to process-based models. They showed the efficiency and generalizability of this paradigm for untrained variables, spatial extrapolation and interpretability. In data-sparse regions, this approach can even produce better daily metrics and future trends than LSTM (Feng et al., 2022a) and can be used to improve flood routing (Bindas et al., 2022). These models seek to combine the power of both ML and physics and have the potential to alleviate data demand, extrapolate better in space and for more extreme conditions, and be constrained by multivariate observations to enable better forecasts. Furthermore, they provide a systematic pathway for asking scientific questions and getting answers from big data.

Explainability is sometimes useful to help develop trust in model predictions. Forecasting agencies frequently engage in a form of story-telling, both for internal and external communications. One reason for providing explainable predictions is that when the forecasts evolve for a given variable, such as spring runoff, users often wish to understand why (i.e. what has changed in the predictors or other factors to explain the change in the predictions). One way to achieve explainability is by providing storylines or narratives around the hybrid forecasts which demonstrate the geophysical credibility of the results. Differentiable modelling can also provide diverse physical variable outputs, trained or untrained, which help develop a narrative (Feng et al., 2022b). Fleming et al. (2021) showed how hydroclimatic storylines can be produced for clients to make the forecast interpretable in terms of understandable geophysical processes. They used pragmatic methods such as 'popular votes' for the candidate predictors cast by a genetic algorithm. The approach revealed how the values of predictors such as antecedent flow and snow water equivalent could help explain the ensemble mean predicted volume. However, there are also limitations to such approaches. Although narratives may help with stakeholder acceptance of hybrid forecasting systems, they can also form a constraint on the forecasting approach, by enforcing consistency of a given prediction method.

4.2 Assimilating human influences

Another emerging challenge is assimilating human influences on the water cycle to obtain better predictions of hydroclimate variables, especially droughts (Brunner et al., 2021; Van Loon et al., 2022). Limited data exist on human impacts such as water storage, groundwater depletion, irrigation, land cover changes, and water transfers. Therefore, how can human decisions, such as the management of reservoir levels or flow abstraction, be integrated within hydrological forecasts? This question is especially relevant over longer timescales, as well as for hydrological forecasting in general, as access to such data is limited (e.g. only very limited information on reservoir operations is included in GloFAS). One option is to develop proxies to detect and model human influence. For instance, census information on the number of households has be used to extend UK urbanisation records (Han et al., 2022). Population density data has also been used as a proxy for urbanisation, to assess the extent to which seasonal streamflow predictability might benefit from 'anthropogenic' predictors such as land cover change alongside seasonal climate forecasts (Slater and Villarini, 2018). (López and Francés, 2013) supplied a dynamic reservoir index alongside climate indices to predict historical annual maximum peak discharge in Spanish rivers. In a large-scale study it was found that reservoir operations could be implicitly simulated by ML approaches that learn from past operations (Ouyang et al., 2021). Lastly, information on the day of the week and on local festivities has been used successfully as a proxy for difference in energy demand (Bogner et al., 2019). Such proxies might also inform a hybrid system on hydro-peaking in rivers downstream from dams.

The lack of accurate predictions of future human activities at the catchment scale is also a major limitation for hydrological forecasting over longer timescales. Here, the increasing coverage and resolution of satellite data may help to provide relevant inputs to hybrid forecasting models such as future predictions of land use change (e.g. Moulds et al., 2015). Emerging satellite altimetry products (e.g. SWOT) may enable a better understanding of reservoir operations, which can be used to constrain hydrological forecasts. Similarly, ML could potentially be used to translate major socio-economic drivers into land cover change. Overall, we suggest that the main bottleneck to integrating human activities in hybrid forecasting systems is not the model algorithms, which can be adapted to any potential predictors, but rather the lack of consistent historical and future time series data on these activities. Unfortunately, this is likely to be a vexing challenge for automated representation. In many reservoir systems, for instance, operations are determined through unpredictable human interactions and negotiations, and may depend on time-varying legal, institutional, ecological and economic factors, such as agricultural markets influencing irrigation practice, or fisheries health directing environmental releases.

4.3 Developing predictive skill

Dynamical forecasts and predictions tend to have low skill over long lead times. The skill of short-term hydroclimatological forecasts is constrained by the skill of meteorological forecasts, which is currently in the range of 3 to 10 days ahead but has been advancing by about one day per decade, such that 'today's 6-day forecast is as accurate as the 5-day forecast ten years ago' (Bauer et al., 2015, p.47). Low flows may have skill up to 20 days in the case of Fundel et al. (2013) and even longer in other cases, especially with good information on initial conditions and/or the memory effect of catchment storage. Seasonal

climate forecasts also have low predictive skill beyond a couple of months, while both seasonal and decadal predictions suffer from the underestimation of atmospheric circulation in climate models, a phenomenon known as the 'signal-to-noise paradox' (e.g. Smith et al., 2020).

One of the advantages of hybrid predictions is that the data-driven methods can be used to enhance predictive skill of the dynamical meteorological or climate forecasts. For instance, decadal predictions are skillful over multiyear forecast periods but have too much uncertainty to provide useful information on interannual variability. Although the CMIP5-6 models can skillfully reproduce certain large-scale circulation patterns, the magnitude of teleconnections tends to be underestimated. Statistical approaches such as 'NAO-matching' attempt to resolve this by selecting members based on their ability to reproduce climate indices and their teleconnections (Smith et al., 2020). Such methods have been employed to enhance decadal streamflow prediction (Moulds et al., 2023) and condition seasonal hydrological forecasts (Donegan et al., 2021). However, further work is still needed to interpret multiyear forecasts to provide actionable information. Given a skillful multiyear forecast, it should be possible to estimate the increased flood or drought risk (for instance) in each year of the forecast period. Data-driven techniques may aid in future developments by trying to draw out the climate model members that perform well in given months or lead times (e.g. Slater et al., 2017).

4.4 Seamless forecasting: merging forecasts, predictions and projections

The utility of hybrid models for 'seamless' hydroclimatic prediction systems spanning weeks to decades is an open research question (Figure 5). There is a growing need for reliable long-term predictions of climate change impacts on the risk of floods and droughts over the coming decades (i.e. 1-40 years ahead), yet reliable information does not exist over such timescales. The lack of seamless climate information is explained by the fact that different scientific weather and climate products have been developed for different applications. Short-term predictions (less than 5 years ahead) tend to rely more on correct initial conditions while long-term predictions and projections (>10 years ahead) rely more on correct external forcings such as greenhouse gases (Boer et al., 2016).

One way to provide longer-term climate impacts information over the coming decades is to constrain uninitialized climate model projections (e.g. climate simulations for the RCP4.5 or RCP8.5 scenarios) using initialized decadal predictions (such as the CMIP6 decadal hindcasts), which tend to better reflect observed climate variability. Befort et al. (2020) developed a method that does this by selecting the climate projections that best match the mean of the decadal predictions over the next 10 years. They showed that the constrained ensemble, which consisted of uninitialized projections for the upcoming 50 years, had higher skill than the full projection ensemble, even after the 10-year period, once decadal prediction information was no longer available. A hybrid system for enhanced prediction of hydroclimatic impacts (e.g. flood risk) could integrate the outputs of such a constrained ensemble.

Beyond the use of uninitialized projections by themselves (covering the whole 1-50 year period), temporally concatenating bias-corrected time series of decadal climate predictions and climate projections is also possible. Befort et al. (2022) assessed different types of bias correction and found that the variance inflation (VINF) method could reduce inconsistencies between the decadal and century-scale time series, especially for central quantiles of the climate time series (close to the multi-model en-

semble median). However, the method could not eliminate all inconsistencies, notably those for extreme quantiles. A seamless hybrid method would therefore be more difficult to generate for hydroclimate extremes such as floods and droughts. However, these two papers (Befort et al., 2020, 2022) open the way for novel research on the merging of decadal predictions and uninitialized projections as input to seamless prediction schemes for hydroclimate impacts using hybrid ML-based approaches.

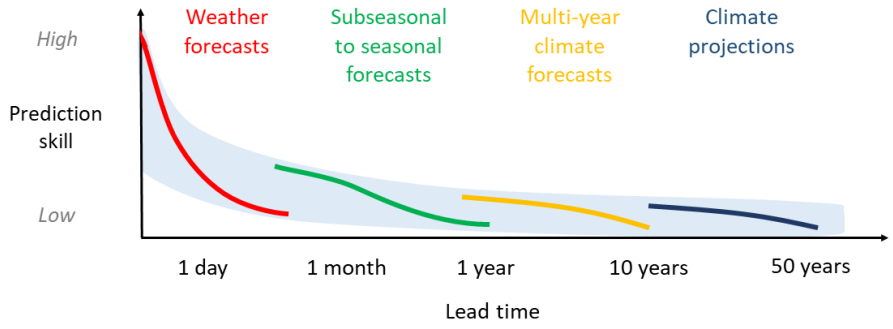

Figure 5. Hybrid models could be a promising route for seamlessly linking initialized predictions from seasonal and decadal forecasts to scenario-based projections across timescales. Different ML-based bias-correction approaches could be explored for merging or concatenating the covariate time series (e.g. Befort et al., 2022) before using them to drive a hybrid hydroclimate prediction model (e.g. for streamflow). Such an approach is likely to be more challenging for extremes such as floods and droughts, and remains an open research question.

4.5 Incorporating spatial variability

The data employed in many hybrid hydrological models are often lumped, i.e. spatially-averaged at the catchment scale, ignoring spatial variability in landscape and atmospheric forcing. Lumped models are challenging for the prediction of hydroclimate in complex environments such as snow-dominated watersheds, which may have karst conduits, or spatiotemporal variation in snow accumulation and snowmelt processes. However, new approaches exist to overcome this limitation in statistical/machine learning models. For instance, Shi et al. (2015) developed a convolutional LSTM, termed convLSTM, which is able to capture spatiotemporal correlations, considering both the input and the prediction target as spatiotemporal sequences. One example is the use of past and future radar maps as input and output: such spatiotemporal sequences have high dimensionality and until recently could not be included in hydroclimate prediction schemes. Similarly, Gupta et al. (2021) developed a spatial variability aware neural network, termed SVANN-E, in which the architecture of the neural network varied spatially across geographic locations. They evaluated the approach using high resolution imagery for wetland mapping. Such novel spatiotemporal prediction approaches are just starting to be used for hydroclimate prediction. Xu et al. (2022) used a hybrid approach to predict streamflow in a watershed with spatially variable karst carbonate bedrock. They combine a spatially-distributed snow model with a deep learning karst model based on convLSTM, which simulated the effect of surface and subsurface properties on the streamflow. This approach allowed the authors to better include the spatial variability in the input variables to their prediction scheme.

4.6 Interpretability, usability, and uptake of hybrid forecasts

Hybrid approaches for hydroclimate prediction over sub-seasonal to decadal lead times face several challenges to their continued uptake by various communities. One issue that is critical to making hybrid schemes more widely accepted is determining whether the improvement in forecast skill obtained by building a hybrid model is worth the extra effort. In other words, it can be difficult to determine *a priori* how much added value can be obtained without first developing the hybrid model and benchmarking the results against a more traditional approach. Despite a commitment to develop the use of ML within operational hydrology (e.g. Environment Agency, 2022), close co-operation is needed between the hydrology, forecasting and ML communities to explore their potential either alone or in hybrid frameworks (Mosavi et al., 2018), build trust (Haupt et al., 2022), communicate skill (Thielen-del Pozo and Bruen, 2019), and overcome barriers to operational uptake (Speight et al., 2021). The benchmarking study of Mai et al. (2022) provided a detailed intercomparison of modelling approaches over the Great Lakes region (USA and Canada), suggesting that the effort related to ML is justifiable. However, this work was for retrospective simulation, rather than forecasting (for which there are more steps needed) and therefore it is still a jump to suggest that ML always provides improvements for prediction, particularly over seasonal to decadal horizons, for which studies are lacking. In the hybrid set-up of Humphrey et al. (2016), for instance, which required the development of both an ML and a conceptual model for three gauges in southern Australia, the authors found that the hybrid model was more skillful than either the conceptual or the data-driven models alone. However, the increase in skill was only marginal for one of the three study locations. They concluded that for this given station, the extra time and effort required to implement the hybrid model was not worth the small gains. Implementing an operational hybrid framework for hydroclimatic forecasting often requires extensive time and expertise, given that two completely different types of models must be developed in parallel. These requirements would also likely require a shift in the expertise of the organisation as well as an upgrade in the computing architecture in the case of GPU-requiring hybrid and data-driven approaches. Overall, the operational uptake of hybrid models is expected to be faster in cases where there is no existing forecasting capability (requiring modification) or where complex physical processes make traditional approaches challenging.

5 Conclusions and remaining research areas in hybrid forecasting

Hybrid forecasting is emerging as a powerful enhancement to traditional hydroclimatic forecasting techniques, but important questions remain regarding their place in the pantheon of methods. We lay out some of the most important research possibilities. First are questions about the evaluation of hybrid methods. How well do dynamical-statistical methods perform when compared with more traditional, operational approaches? What benchmarks should be used? How reliable are these models, and over what lead times can they be trusted? As far as we are aware, there have been very few papers (if any) comparing the skill of hybrid models with operational systems. One systematic comparison of 13 different models (including machine-learning-based, basin-wise, subbasin-based, and gridded models) revealed the superiority of the data-driven LSTM-lumped model in all experiments (e.g. Mai et al., 2022), suggesting that hybrid LSTM-based prediction systems would be a promising route for daily simulation, and potentially for applications such as forecasting.

Second are questions about the potential for seamless prediction. To what extent can hybrid approaches be employed to meld historical trends, near-term and decadal predictions of hydroclimate variables from atmospheric forecasts, climate model predictions, and projections? How would such a system be used operationally? Seamless hybrid prediction may provide better insights into long-term hydroclimatic trends, but merging across time-scales can lead to inconsistencies in the time series (i.e. 'jumps' or step-changes) between e.g. decadal climate predictions and the climate projections (Befort et al., 2022). Third are questions about use of data-driven models to detect and attribute the drivers of hydrologic change (Slater et al., 2021), and then integrate such knowledge within a predictive framework. How can data-driven approaches be employed to understand the relative contributions of different predictors, including human impacts such as the effects of reservoirs on streamflow (Brunner and Naveau, 2022)? To what extent can hybrid models uncover 'hidden' large-scale climatic or anthropogenic drivers of change (Renard et al., 2022; Lees et al., 2022)?

An important step forward would be the development of consistent global datasets of climate hindcasts at various time scales at the catchment level. Similar datasets developed for large sample hydrological analyses such as CAMELS (e.g. Addor et al., 2017; Coxon et al., 2020) and Caravan (Kratzert et al., 2023) have driven rapid progress in ML methods for simulating daily streamflow using observed climate inputs. Such datasets drive progress towards operational hybrid systems by making it easier for model developers to train and test potential methods in a pseudo-operational context. Moreover, they could integrate consistent estimates of other potential drivers – including streamflow signatures and local characteristics related to topography, geology and land cover (as in the CAMELS datasets) – enabling forecasters to understand the contribution of different drivers to streamflow predictability across time scales.

Finally, there are questions about the acceptance and viability of hybrid models in operational contexts, given the dominance, familiarity with and deep embedding of physics-based forecasting and prediction methods (Cohen et al., 2019). In what ways could hybrid approaches complement, support, or replace conventional physically-based systems? The pace of change in such settings is often constrained by practicalities, institutional resistance (Arnal et al., 2020) or the requirement for decision-relevant evidence of skill. Acceptance might be advanced by systematically comparing the outputs from hybrid models with operational models under identical forcings, to assess the physical interpretation of model results (e.g. Mai et al., 2022). To convince operational forecasters that hybrid models may add value alongside more traditional approaches requires rigorous benchmarking by the community alongside established approaches. It may also require more extensive changes in the education and preparation of the workforce that is needed to staff operational centres.

There are several possible paths forward. One of these frames hybrid models not as replacing current operational systems but as a complementary tool, extension or enhancement, helping on different levels, and likely within existing systems. Another path forward is to recognize the difference in skill between hybrid models vs. traditional models, and to start to develop future replacements for current operational models; replacements based fundamentally on data-driven (ML, DL, even AI) principles, but with the ability to incorporate elements of traditional hydrological and climate science where these are beneficial. Furthermore, hybrid models could be developed to estimate both impacts and mitigation measures, based on past events. All these approaches make sense for different reasons and in different scenarios, and various agencies and organizations are pursuing both these and other strategies for incorporating data-driven methods into operational workflows. Overall, the utility

of hybrid models is not only for enhancing forecasting and prediction, but also for allowing deeper interrogation of diverse data, revealing sometimes hidden or obscure hydroclimatological processes.

Author contributions. LJS led the review and all other authors contributed equally to writing the paper.

Code availability. There is no code or data associated with this review article.

Competing interests. At least one of the authors is a member of the editorial board of Hydrology and Earth System Sciences.

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W., Love, C., Madadgar, S., Papalexiou, S., Davis, S., et al.: Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting, Philosophical Transactions of the Royal Society A, 380, 20210 288, 2022.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS–global ensemble streamflow forecasting and flood early warning, Hydrology and Earth System Sciences, 17, 1161–1175, 2013.

Altman, N. and Krzywinski, M.: The curse(s) of dimensionality, Nature Methods, 15, 399–400, 2018.

Anctil, F., Michel, C., Perrin, C., and Andréassian, V.: A soil moisture index as an auxiliary ANN input for stream flow forecasting, Journal of Hydrology, 286, 155–167, 2004.

Anderson, G. J. and Lucas, D. D.: Machine Learning Predictions of a Multiresolution Climate Model Ensemble, Geophysical Research Letters, 45, 4273–4280, https://doi.org/10.1029/2018GL077049, 2018.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, Hydrology and Earth System Sciences, 24, 535–559, 2020.

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B., and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, Hydrology and Earth System Sciences, 22, 2057–2072, 2018.

Arnal, L., Anspoks, L., Manson, S., Neumann, J., Norton, T., Stephens, E., Wolfenden, L., and Cloke, H. L.: "Are we talking just a bit of water out of bank? Or is it Armageddon?" Front line perspectives on transitioning to probabilistic fluvial flood forecasts in England, Geoscience Communication, 3, 203–232, 2020.

Baker, S., Wood, A., and Rajagopalan, B.: Application of Postprocessing to Watershed-Scale Subseasonal Climate Forecasts over the Contiguous United States, Journal of Hydrometeorology, 21, 971 – 987, https://doi.org/10.1175/JHM-D-19-0155.1, 2020.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, 2015.

Befort, D., Brunner, L., Borchert, L., O'Reilly, C., Mignot, J., Ballinger, A., Hegerl, G., Murphy, J., and Weisheimer, A.: Combination of decadal predictions and climate projections in time: Challenges and potential solutions, Geophysical Research Letters, p. e2022GL098568, 2022.

Befort, D. J., O'Reilly, C. H., and Weisheimer, A.: Constraining projections using decadal predictions, Geophysical Research Letters, 47, e2020GL087 900, 2020.

Bennett, J. C., Wang, Q., Li, M., Robertson, D. E., and Schepen, A.: Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model, Water Resources Research, 52, 8238–8259, 2016.

Bennett, J. C., Robertson, D. E., Wang, Q. J., Li, M., and Perraud, J.-M.: Propagating reliable estimates of hydrological forecast uncertainty to many lead times, Journal of Hydrology, 603, 126 798, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126798, 2021a.

Bennett, J. C., Wang, Q., Robertson, D. E., Bridgart, R., Lerat, J., Li, M., and Michael, K.: An error model for long-range ensemble forecasts of ephemeral rivers, Advances in Water Resources, 151, 103 891, 2021b.

Bergström, S.: Development and application of a conceptual model for Scandinavian catchments, Tech. Rep. Report RHO No. 7, Norrköping, Sweden, 1976.

Beven, K.: Deep learning, hydrological processes and the uniqueness of place, Hydrological Processes, 34, 3608–3613, 2020.

Bindas, T., Tsai, W.-P., Liu, J., Rahmani, F., Feng, D., Bian, Y., Lawson, K., and Shen, C.: Improving large-basin streamflow simulation using a modular, differentiable, learnable graph model for routing, Authorea Preprints, 2022.

Bisson, J. and Roberge, F.: Prévisions des apports naturels: Expérience d'Hydro-Québec, in: Workshop on flow predictions, Toronto, 1983.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., et al.: The decadal climate prediction project (DCPP) contribution to CMIP6, Geoscientific Model Development, 9, 3751–3777, 2016.

Bogner, K., Pappenberger, F., and Zappa, M.: Machine Learning Techniques for Predicting the Energy Consumption/Production and Its Uncertainties Driven by Meteorological Observations and Forecasts, Sustainability, 11, 3328, https://doi.org/10.3390/su11123328, 2019.

Bogner, K., Chang, A. Y., Bernhard, L., Zappa, M., Monhart, S., and Spirig, C.: Tercile Forecasts for Extending the Horizon of Skillful Hydrological Predictions, Journal of Hydrometeorology, 23, 521–539, https://doi.org/10.1175/JHM-D-21-0020.1, 2022.

Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., Watt-Meyer, O., McGibbon, J., Perkins, W. A., Clark, S. K., and Harris, L.: Correcting Coarse-Grid Weather and Climate Models by Machine Learning From Global Storm-Resolving Simulations, Journal of Advances in Modeling Earth Systems, 14, https://doi.org/10.1029/2021MS002794, 2022.

Brunner, M. I. and Naveau, P.: Disentangling natural streamflow from reservoir regulation practices in the Alps using generalized additive models, Hydrology and Earth System Sciences Discussions, 2022, 1–17, https://doi.org/10.5194/hess-2022-244, 2022.

Brunner, M. I., Slater, L., Tallaksen, L. M., and Clark, M.: Challenges in modeling and predicting floods and droughts: A review, Wiley Interdisciplinary Reviews: Water, p. e1520, 2021.

Burnash, R. J., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system: Conceptual modeling for digital computers, US Department of Commerce, National Weather Service, and State of California . . . , 1973.

Cao, J., Wang, H., Li, J., Tian, Q., and Niyogi, D.: Improving the Forecasting of Winter Wheat Yields in Northern China with Machine Learning–Dynamical Hybrid Subseasonal-to-Seasonal Ensemble Prediction, Remote Sensing, 14, 1707, 2022.

Chang, A. Y., Bogner, K., Grams, C. M., Monhart, S., Domeisen, D. I., and Zappa, M.: Exploring the use of European weather regimes for improving user-relevant hydrological forecasts at the sub-seasonal scale in Switzerland, Journal of Hydrometeorology, 2022, under review.

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., and Tziperman, E.: S2S Reboot: An Argument for Greater Inclusion of Machine Learning in Subseasonal to Seasonal Forecasts, WIREs Climate Change, 10, https://doi.org/10.1002/wcc.567, 2019.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J., Lane, R., Lewis, M., Robinson, E. L., et al.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth System Science Data, 12, 2459–2483, 2020.

Crawford, N. and Thurin, S.: Hydrologic estimates for small hydroelectric projects, Tech. rep., Washington, DC, USA, 1981.

DelSole, T. and Shukla, J.: Artificial skill due to predictor screening, Journal of Climate, 22, 331–345, 2009.

Dixon, S. G. and Wilby, R. L.: A seasonal forecasting procedure for reservoir inflows in Central Asia, River Research and Applications, 35, 1141–1154, 2019.

Donegan, S., Murphy, C., Harrigan, S., Broderick, C., Foran Quinn, D., Golian, S., Knight, J., Matthews, T., Prudhomme, C., Scaife, A. A., et al.: Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times, Hydrology and Earth System Sciences, 25, 4159–4183, 2021.

Duan, Q., Pappenberger, F., Wood, A., Cloke, H. L., and Schaake, J.: Handbook of hydrometeorological ensemble forecasting, vol. 845, Springer Berlin/Heidelberg, Germany, 2019.

Duan, S., Ullrich, P., and Shu, L.: Using convolutional neural networks for streamflow projection in california, Frontiers in Water, 2, 28, 2020.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P., and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geoscientific Model Development, 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Environment Agency: Flood Hydrology Roadmap: Roadmap development and the action plan (FRS18196/R1), Tech. rep., https://assets.publishing.service.gov.uk/media/62335ac2e90e070a54e18185/FRS18196_Flood_hydrology_roadmap_-_report.pdf, 2022.

Essenfelder, A. H., Larosa, F., Mazzoli, P., Bagli, S., Broccoli, D., Luzzi, V., Mysiak, J., Mercogliano, P., and dalla Valle, F.: Smart Climate Hydropower Tool: A Machine-Learning Seasonal Forecasting Climate Service to Support Cost–Benefit Analysis of Reservoir Management, Atmosphere, 11, 1305, 2020.

Fang, K. and Shen, C.: Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel, Journal of Hydrometeorology, 21, 399 – 413, https://doi.org/10.1175/JHM-D-19-0169.1, 2020a.

Fang, K. and Shen, C.: Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel, Journal of Hydrometeorology, 21, 399–413, 2020b.

Fang, K., Shen, C., Kifer, D., and Yang, X.: Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network, Geophysical Research Letters, 44, 11 030–11 039, 2017.

Fang, K., Kifer, D., Lawson, K., Feng, D., and Shen, C.: The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology, Water Resources Research, 58, e2021WR029 583, https://doi.org/https://doi.org/10.1029/2021WR029583, e2021WR029583 2021WR029583, 2022.

Feng, D., Fang, K., and Shen, C.: Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales, Water Resources Research, 56, e2019WR026 793, 2020.

Feng, D., Lawson, K., and Shen, C.: Mitigating Prediction Error of Deep Learning Streamflow Models in Large Data-Sparse Regions With Ensemble Modeling and Soft Data, Geophysical Research Letters, 48, e2021GL092 999, https://doi.org/https://doi.org/10.1029/2021GL092999, e2021GL092999 2021GL092999, 2021.

Feng, D., Beck, H., Lawson, K., and Shen, C.: The suitability of differentiable, learnable hydrologic models for ungauged regions and climate change impact assessment, Hydrology and Earth System Sciences Discussions, 2022, 1–28, https://doi.org/10.5194/hess-2022-245, 2022a.

Feng, D., Liu, J., Lawson, K., and Shen, C.: Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy, Water Resources Research, 58, e2022WR032 404, 2022b.

Fleming, S. W., Garen, D. C., Goodbody, A. G., McCarthy, C. S., and Landers, L. C.: Assessing the new Natural Resources Conservation Service water supply forecast model for the American West: A challenging test of explainable, automated, ensemble artificial intelligence, Journal of Hydrology, 602, 126 782, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126782, 2021.

Flora, M. L., Potvin, C. K., Skinner, P. S., Handler, S., and McGovern, A.: Using Machine Learning to Generate Storm-Scale Probabilistic Guidance of Severe Weather Hazards in the Warn-on-Forecast System, Monthly Weather Review, 149, 1535 – 1557, https://doi.org/10.1175/MWR-D-20-0194.1, 2021.

Frame, J., Kratzert, F., Klotz, D., Gauch, M., Shelev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S.: Deep learning rainfall-runoff predictions of extreme events, Hydrology and Earth System Sciences, 26, 3377–3392, 2022a.

Frame, J., Ullrich, P., Nearing, G., Gupta, H., and Kratzert, F.: On strictly enforced mass conservation constraints for modeling the rainfall-runoff process, 2022b.

Frame, J. M., Kratzert, F., Raney, A., Rahman, M., Salas, F. R., and Nearing, G. S.: Post-Processing the National Water Model with Long Short-Term Memory Networks for Streamflow Predictions and Model Diagnostics, JAWRA Journal of the American Water Resources Association, 57, 885–905, 2021.

Freeze, R. A. and Harlan, R.: Blueprint for a physically-based, digitally-simulated hydrologic response model, Journal of hydrology, 9, 237–258, 1969.

Fundel, F., Jörg-Hess, S., and Zappa, M.: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices, Hydrology and Earth System Sciences, 17, 395–407, 2013.

Garen, D. C.: Improved techniques in regression-based streamflow volume forecasting, Journal of Water Resources Planning and Management, 118, 654–670, 1992.

Gibson, P. B., Chapman, W. E., Altinok, A., Delle Monache, L., DeFlorio, M. J., and Waliser, D. E.: Training Machine Learning Models on Climate Model Output Yields Skillful Interpretable Seasonal Precipitation Forecasts, Communications Earth & Environment, 2, 159, https://doi.org/10.1038/s43247-021-00225-4, 2021.

Glahn, H. R. and Lowry, D. A.: The use of model output statistics (MOS) in objective weather forecasting, Journal of Applied Meteorology and Climatology, 11, 1203–1211, 1972.

Golian, S., Murphy, C., and Meresa, H.: Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland, Journal of Hydrology: Regional Studies, 36, 100 859, 2021.

Golian, S., Murphy, C., Wilby, R. L., Matthews, T., Donegan, S., Quinn, D. F., and Harrigan, S.: Dynamical-statistical seasonal forecasts of winter and summer precipitation for the Island of Ireland, International Journal of Climatology, 42, 5714–5731, 2022.

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S., and Hoefler, T.: Deep Learning for Post-Processing Ensemble Weather Forecasts, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379, 20200 092, https://doi.org/10.1098/rsta.2020.0092, 2021.

Gupta, J., Molnar, C., Xie, Y., Knight, J., and Shekhar, S.: Spatial Variability Aware Deep Neural Networks (SVANN): A General Approach, ACM Trans. Intell. Syst. Technol., 12, https://doi.org/10.1145/3466688, 2021.

Hagen, J. S., Leblois, E., Lawrence, D., Solomatine, D., and Sorteberg, A.: Identifying major drivers of daily streamflow from large-scale atmospheric circulation with machine learning, Journal of Hydrology, 596, 126 086, 2021.

Han, S., Slater, L., Wilby, R. L., and Faulkner, D.: Contribution of Urbanisation to Non-stationary River Flow in the UK, Journal of Hydrology, p. 128417, 2022.

Hapuarachchi, H., Bari, M., Kabir, A., Hasan, M., Woldemeskel, F., Gamage, N., Sunter, P., Zhang, X., Robertson, D., Bennett, J., et al.: Development of a national 7-day ensemble streamflow forecasting service for Australia, Hydrology and Earth System Sciences Discussions, pp. 1–35, 2022.

Harrigan, S., Zoster, E., Cloke, H., Salamon, P., and Prudhomme, C.: Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System, Hydrology and Earth System Sciences Discussions, 2020, 1–22, https://doi.org/10.5194/hess-2020-532, 2020.

Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., and Palmer, T. N.: A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts, arXiv preprint arXiv:2204.02028, 2022.

Haupt, S. E., Gagne, D. J., Hsieh, W. W., Krasnopolsky, V., McGovern, A., Marzban, C., Moninger, W., Lakshmanan, V., Tissot, P., and Williams, J. K.: The History and Practice of AI in the Environmental Sciences, Bulletin of the American Meteorological Society, 103, E1351 – E1370, https://doi.org/10.1175/BAMS-D-20-0234.1, 2022.

Hauswirth, S. M., Bierkens, M. F., Beijk, V., and Wanders, N.: The suitability of a hybrid framework including data driven approaches for hydrological forecasting, Hydrology and Earth System Sciences Discussions, pp. 1–20, 2022.

Hemri, S., Fundel, F., and Zappa, M.: Simultaneous calibration of ensemble river flow predictions over an entire range of lead times, Water Resources Research, 49, 6744–6755, https://doi.org/10.1002/wrcr.20542, 2013.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, Journal of Hydrology, 566, 595–606, 2018.

Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G. S., Hochreiter, S., and Klambauer, G.: Mc-lstm: Mass-conserving LSTM, in: International Conference on Machine Learning, pp. 4275–4286, PMLR, 2021.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., and Maier, H. R.: A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network, Journal of Hydrology, 540, 623–640, 2016.

Jain, S. K., Mani, P., Jain, S. K., Prakash, P., Singh, V. P., Tullos, D., Kumar, S., Agarwal, S., and Dimri, A.: A Brief review of flood forecasting techniques and their applications, International Journal of River Basin Management, 16, 329–344, 2018.

Jörg-Hess, S., Griessinger, N., and Zappa, M.: Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas, Journal of Hydrometeorology, 16, 2169 – 2186, https://doi.org/10.1175/JHM-D-14-0193.1, 2015.

Kang, N. and Elsner, J. B.: Interpretation of the statistical/dynamical prediction for seasonal tropical storm frequency in the western North Pacific, Environmental Research Letters, 16, 014 017, https://doi.org/10.1088/1748-9326/abcdd3, 2020.

Khouakhi, A., Villarini, G., Zhang, W., and Slater, L. J.: Seasonal predictability of high sea level frequency using ENSO patterns along the US West Coast, Advances in Water Resources, 131, 103 377, 2019.

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., Becker, E., et al.: The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, Bulletin of the American Meteorological Society, 95, 585–601, 2014.

Klotzbach, P., Caron, L.-P., and Bell, M.: A statistical/dynamical model for North Atlantic seasonal hurricane prediction, Geophysical Research Letters, 47, e2020GL089 357, 2020.

Krabbenhoft, C. A.: Assessing placement bias of the global river gauge network, Nature Sustainability, 5, 10, https://doi.org/10.1038/s41893-022-00873-0, 2022.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, 55, 11 344–11 354, 2019a.

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

Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, Hydrology and Earth System Sciences, 25, 2685–2703, 2021.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., et al.: Caravan-A global community dataset for large-sample hydrology, Scientific Data, 10, 61, 2023.

Kumanlioglu, A. A. and Fistikoglu, O.: Performance enhancement of a conceptual hydrological model by integrating artificial intelligence, Journal of Hydrologic Engineering, 24, 04019 047, 2019.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J.: Benchmarking Data-Driven Rainfall-Runoff Models in Great Britain: A comparison of LSTM-based models with four lumped conceptual models, Hydrology and Earth System Sciences, 25, 5517–5534, 2021.

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., and Dadson, S.: Hydrological concept formation inside long short-term memory (LSTM) networks, Hydrology and Earth System Sciences, 26, 3079–3101, 2022.

Lehner, F., Wood, A. W., Llewellyn, D., Blatchford, D. B., Goodbody, A. G., and Pappenberger, F.: Mitigating the impacts of climate nonstationarity on seasonal streamflow predictability in the US Southwest, Geophysical Research Letters, 44, 12–208, 2017.

Li, Y., Wu, Z., He, H., and Yin, H.: Sub-seasonal precipitation forecasts using preceding atmospheric intraseasonal oscillation signals in a Bayesian perspective, Hydrology and Earth System Sciences Discussions, pp. 1–30, 2022.

Liu, W., Yang, T., Sun, F., Wang, H., Feng, Y., and Du, M.: Observation-constrained projection of global flood magnitudes with anthropogenic warming, Water Resources Research, 57, e2020WR028 830, 2021.

López, J. and Francés, F.: Non-Stationary Flood Frequency Analysis in Continental Spanish Rivers, Using Climate and Reservoir Indices as External Covariates, Hydrology and Earth System Sciences, 17, 3189–3203, https://doi.org/10.5194/hess-17-3189-2013, 2013.

Ma, J., Sun, J., and Liu, C.: A hybrid statistical-dynamical prediction scheme for summer monthly precipitation over Northeast China, Meteorological Applications, 29, e2057, 2022.

Ma, K., Feng, D., Lawson, K., Tsai, W.-P., Liang, C., Huang, X., Sharma, A., and Shen, C.: Transferring Hydrologic Data Across Continents – Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions, Water Resources Research, 57, e2020WR028 600, https://doi.org/https://doi.org/10.1029/2020WR028600, e2020WR028600 2020WR028600, 2021.

Madadgar, S., AghaKouchak, A., Shukla, S., Wood, A. W., Cheng, L., Hsu, K.-L., and Svoboda, M.: A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States, Water Resources Research, 52, 5095–5110, 2016.

Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, R., Fortin, V., Fry, L. M., Gauch, M., Klotz, D., Kratzert, F., O'Brien, N., Princz, D. G., Koya, S. R., Roy, T., Seglenieks, F., Shrestha, K., Temgoua, A. G. T., Vionnet, V., and Waddell, J. W.: The Great Lakes Runoff Intercomparison Project Phase 4: The Great Lakes (GRIP-GL), 26, 54, 2022.

Massoud, E. C., Lee, H., Gibson, P. B., Loikith, P., and Waliser, D. E.: Bayesian Model Averaging of Climate Model Projections Constrained by Precipitation Observations over the Contiguous United States, Journal of Hydrometeorology, 21, 2401–2418, https://doi.org/10.1175/JHM-D-19-0258.1, 2020.

McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Woldemeskel, F., Tuteja, N., and Kuczera, G.: Seamless streamflow model provides forecasts at all scales from daily to monthly and matches the performance of non-seamless monthly model, Hydrology and Earth System Sciences Discussions, 2022, 1–22, https://doi.org/10.5194/hess-2021-589, 2022.

Meißner, D., Klein, B., and Ionita, M.: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, Hydrology and Earth System Sciences, 21, 6401–6423, https://doi.org/10.5194/hess-21-6401-2017, 2017.

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., Brekke, L. D., and Arnold, J. R.: An intercomparison of approaches for improving operational seasonal streamflow forecasts, Hydrology and Earth System Sciences, 21, 3915–3935, 2017.

Met Office, Environment Agency and Flood Forecasting Centre: Talking the Same Language. Updated with learning from the 2012 Floods, https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/business/public-sector/hazard-manager/glossary_for_talking_the_same_language.pdf, 2013.

Miao, Q., Pan, B., Wang, H., Hsu, K., and Sorooshian, S.: Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network, Water, 11, 977, 2019.

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J.: Stationarity is dead: Whither water management?, Science, 319, 573–574, 2008.

Mohammadi, B., Moazenzadeh, R., Christian, K., and Duan, Z.: Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models, Environmental Science and Pollution Research, 28, 65 752–65 768, 2021.

Monhart, S., Zappa, M., Spirig, C., Schär, C., and Bogner, K.: Subseasonal hydrometeorological ensemble predictions in small- and medium-sized mountainous catchments: Benefits of the NWP approach, Hydrology and Earth System Sciences, 23, 493–513, https://doi.org/10.5194/hess-23-493-2019, 2019.

Moon, S.-H., Kim, Y.-H., Lee, Y. H., and Moon, B.-R.: Application of machine learning to an early warning system for very short-term heavy rainfall, Journal of Hydrology, 568, 1042–1054, https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.11.060, 2019.

Mosavi, A., Ozturk, P., and Chau, K.-w.: Flood prediction using machine learning models: Literature review, Water, 10, 1536, 2018.

Moulds, S., Buytaert, W., and Mijic, A.: An Open and Extensible Framework for Spatially Explicit Land Use Change Modelling: the lulcc R package, Geoscientific Model Development, 8, 3215–3229, https://doi.org/10.5194/gmd-8-3215-2015, 2015.

Moulds, S., Slater, Louise Dunstone, N., and Smith, D.: Skillful decadal flood prediction, Geophysical Research Letters, 49, e2022GL100 650, https://doi.org/10.1029/2022GL100650, 2023.

Murakami, H., Villarini, G., Vecchi, G. A., Zhang, W., and Gudgel, R.: Statistical–dynamical seasonal forecast of North Atlantic and US landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model, Monthly Weather Review, 144, 2101–2123, 2016.

Najafi, H., Robertson, A. W., Massah Bavani, A. R., Irannejad, P., Wanders, N., and Wood, E. F.: Improved multi-model ensemble forecasts of Iran's precipitation and temperature using a hybrid dynamical-statistical approach during fall and winter seasons, International Journal of Climatology, 41, 5698–5725, 2021.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

Neal, R., Fereday, D., Crocker, R., and Comer, R. E.: A flexible approach to defining weather patterns and their application in weather forecasting over Europe, Meteorological Applications, 23, 389–400, 2016.

Neal, R., Dankers, R., Saulter, A., Lane, A., Millard, J., Robbins, G., and Price, D.: Use of probabilistic medium-to long-range weather-pattern forecasts for identifying periods with an increased likelihood of coastal flooding around the UK, Meteorological Applications, 25, 534–547, 2018.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What role does hydrological science play in the age of machine learning?, Water Resources Research, 57, e2020WR028 091, 2021.

Nearing, G. S., Klotz, D., Frame, J. M., Gauch, M., Gilon, O., Kratzert, F., Sampson, A. K., Shalev, G., and Nevo, S.: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks, Hydrology and Earth System Sciences, 26, 5493–5513, https://doi.org/10.5194/hess-26-5493-2022, 2022.

Neri, A., Villarini, G., Salvi, K. A., Slater, L. J., and Napolitano, F.: On the decadal predictability of the frequency of flood events across the US Midwest, International Journal of Climatology, 39, 1796–1804, 2019.

Neri, A., Villarini, G., and Napolitano, F.: Intraseasonal predictability of the duration of flooding above National Weather Service flood warning levels across the US Midwest, Hydrological Processes, 34, 4505–4511, 2020.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Peled Levi, N., Reich, O., Gilon, O., Maor, R., Timnat, S., Shechter, T., Anisimov, V., Gigi, Y., Levin, Y., Moshe, Z., Ben-Haim, Z., Hassidim, A., and Matias, Y.: Flood forecasting with machine learning models in an operational framework, Hydrology and Earth System Sciences, 26, 4013–4032, https://doi.org/10.5194/hess-26-4013-2022, 2022.

Newman, A., Clark, M., Sampson, K., Wood, A., Hay, L., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J., et al.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, 2015.

Nilsson, P., Uvo, C., and Berndtsson, R.: Monthly runoff simulation: Comparing and combining conceptual and neural network models., Journal of Hydrology, 321, 344–363, 2006.

NOAA: National Water Model: Improving NOAA's Water Prediction Services, 2016.

Okkan, U., Ersoy, Z., Kumanlioglu, A., and Fistikoglu, O.: Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: A nested hybrid rainfall-runoff modeling, Journal of Hydrology, 598, 2021.

Ouyang, W., Lawson, K., Feng, D., Ye, L., Zhang, C., and Shen, C.: Continental-scale streamflow modeling of basins with reservoirs: Towards a coherent deep-learning-based strategy, Journal of Hydrology, 599, 126 455, 2021.

Pan, B., Anderson, G. J., Goncalves, A., Lucas, D. D., Bonfils, C. J. W., and Lee, J.: Improving Seasonal Forecast Using Probabilistic Deep Learning, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 766, https://doi.org/https://doi.org/10.1029/2021MS002766, e2021MS002766 2021MS002766, 2022.

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., Bell, R., DelSole, T., Min, D., Zhu, Y., et al.: The Subseasonal Experiment (SubX): A multimodel subseasonal prediction experiment, Bulletin of the American Meteorological Society, 100, 2043–2060, 2019.

Perrin, C., Michel, C., and Andréeassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, 2003.

Piadeh, F., Behzadian, K., and Alani, A. M.: A critical review of real-time modelling of flood forecasting in urban drainage systems, Journal of Hydrology, 607, 127 476, https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127476, 2022.

Pilling, C., Dodds, V., Cranston, M., Price, D., Harrison, T., and How, A.: Flood forecasting—A national overview for great britain, in: Flood forecasting, pp. 201–247, Elsevier, 2016.

Rasouli, K., Hsieh, W. W., and Cannon, A. J.: Daily streamflow forecasting by machine learning methods with weather and climate inputs, Journal of Hydrology, 414, 284–293, 2012.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S.: Skilful Precipitation Nowcasting Using Deep Generative Models of Radar, Nature, 597, 672–677, https://doi.org/10.1038/s41586-021-03854-z, 2021.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, 2019.

Ren, W., Yang, T., Huang, C., Xu, C., and Shao, Q.: Improving monthly streamflow prediction in Alpine regions: integrating HBV model with Bayesian neural network, Stochastic Environmental Research and Risk Assessment, 32, 3381–3396, 2018.

Renard, B. and Thyer, M.: Revealing hidden climate indices from the occurrence of hydrologic extremes, Water Resources Research, 55, 7662–7681, 2019.

Renard, B., Thyer, M., McInerney, D., Kavetski, D., Leonard, M., and Westra, S.: A Hidden Climate Indices Modeling Framework for Multivariable Space-Time Data, Water Resources Research, 58, e2021WR030 007, 2022.

Richardson, D., Neal, R., Dankers, R., Mylne, K., Cowling, R., Clements, H., and Millard, J.: Linking weather patterns to regional extreme precipitation for highlighting potential flood events in medium-to long-range forecasts, Meteorological Applications, 27, e1931, 2020.

Rözer, V., Peche, A., Berkhahn, S., Feng, Y., Fuchs, L., Graf, T., Haberlandt, U., Kreibich, H., Sämann, R., Sester, M., et al.: Impact-based forecasting for pluvial floods, Earth's Future, 9, 2020EF001 851, 2021.

Sabeerali, C., Sreejith, O., Acharya, N., Surendran, D. E., and Pai, D.: Seasonal Forecasting of Tropical Cyclones over the Bay of Bengal using a Hybrid Statistical/Dynamical Model, International Journal of Climatology, 2022.

Sahu, N., Robertson, A. W., Boer, R., Behera, S., DeWitt, D. G., Takara, K., Kumar, M., and Singh, R.: Probabilistic seasonal streamflow forecasts of the Citarum River, Indonesia, based on general circulation models, Stochastic Environmental Research and Risk Assessment, 31, 1747–1758, 2017.

Salvi, K., Villarini, G., and Vecchi, G. A.: High resolution decadal precipitation predictions over the continental United States for impacts assessment, Journal of Hydrology, 553, 559–573, 2017a.

Salvi, K., Villarini, G., Vecchi, G. A., and Ghosh, S.: Decadal temperature predictions over the continental United States: Analysis and Enhancement, Climate dynamics, 49, 3587–3604, 2017b.

Scher, S., Jewson, S., and Messori, G.: Robust Worst-Case Scenarios from Ensemble Forecasts, Weather and Forecasting, 36, 1357 – 1373, https://doi.org/10.1175/WAF-D-20-0219.1, 2021.

Schick, S., Rössler, O., and Weingartner, R.: Monthly streamflow forecasting at varying spatial scales in the Rhine basin, Hydrology and Earth System Sciences, 22, 929–942, https://doi.org/10.5194/hess-22-929-2018, 2018.

Schlef, K. E., François, B., and Brown, C.: Comparing flood projection approaches across hydro-climatologically diverse United States river basins, Water Resources Research, 57, e2019WR025 861, 2021.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., et al.: Differentiable modeling to unify machine learning and physical models and advance Geosciences, arXiv preprint arXiv:2301.04027, 2023.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, Advances in neural information processing systems, 28, https://proceedings.neurips.cc/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf, 2015.

Sivapalan, M.: Prediction in Ungauged Basins: A Grand Challenge for Theoretical Hydrology, Hydrological Processes, 17, 3163–3170, https://doi.org/10.1002/hyp.5155, 2003.

Slater, L. J. and Villarini, G.: Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach, Geophysical Research Letters, 45, 6504–6513, 2018.

Slater, L. J., Villarini, G., and Bradley, A. A.: Weighting of NMME temperature and precipitation forecasts across Europe, Journal of Hydrology, 552, 646–659, 2017.

Slater, L. J., Villarini, G., Bradley, A. A., and Vecchi, G. A.: A dynamical statistical framework for seasonal streamflow forecasting in an agricultural watershed, Climate Dynamics, 53, 7429–7445, 2019.

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., et al.: Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management, Hydrology and Earth System Sciences, 25, 3897–3935, 2021.

Slater, L. J., Huntingford, C., Pywell, R. F., Redhead, J. W., and Kendon, E. J.: Resilience of UK crop yields to compound climate change, Earth System Dynamics, 13, 1377–1396, 2022.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Müller, W. A., Nicolí, D., Ortega, P., Pankatz,

K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic Climate Far More Predictable than Models Imply, Nature, 583, 796–800, https://doi.org/10.1038/s41586-020-2525-0, 2020.

Smith, P., Pappenberger, F., Wetterhall, F., Thielen del Pozo, J., Krzeminski, B., Salamon, P., Muraro, D., Kalas, M., and Baugh, C.: Chapter 11 - On the Operational Implementation of the European Flood Awareness System (EFAS), in: Flood Forecasting, edited by Adams, T. E. and Pagano, T. C., pp. 313–348, Academic Press, Boston, https://doi.org/https://doi.org/10.1016/B978-0-12-801884-2.00011-6, 2016.

Speight, L. J., Cranston, M. D., White, C. J., and Kelly, L.: Operational and emerging capabilities for surface water flood forecasting, Wiley Interdisciplinary Reviews: Water, 8, e1517, 2021.

Thielen, J., Bartholmes, J., Ramos, M.-H., and De Roo, A.: The European flood alert system–Part 1: concept and development, Hydrology and Earth System Sciences, 13, 125–140, 2009.

Thielen-del Pozo, J. and Bruen, M.: Overview of forecast communication and use of ensemble hydrometeorological forecasts, Handbook of Hydrometeorological Ensemble Forecasting, pp. 1037–1045, 2019.

Tian, D., He, X., Srivastava, P., and Kalin, L.: A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information, Stochastic Environmental Research and Risk Assessment, pp. 1–23, 2021.

Toms, B. A., Barnes, E. A., and Ebert-Uphoff, I.: Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability, Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019MS002002, 2020.

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years, https://doi.org/https://doi.org/10.1029/2020WR028392, 2021.

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., and Shen, C.: From Calibration to Parameter Learning: Harnessing the Scaling Effects of Big Data in Geoscientific Modeling, Nature Communications, 12, 5988, https://doi.org/10.1038/s41467-021-26107-z, 2021.

Unger, D. A., van den Dool, H., O'Lenic, E., and Collins, D.: Ensemble regression, Monthly Weather Review, 137, 2365–2379, 2009.

Van Loon, A. F., Rangecroft, S., Coxon, G., Werner, M., Wanders, N., Di Baldassarre, G., Tijdeman, E., Bosman, M., Gleeson, T., Nauditt, A., et al.: Streamflow droughts aggravated by human activities despite management, Environmental Research Letters, 17, 044 059, 2022.

Vecchi, G. A., Zhao, M., Wang, H., Villarini, G., Rosati, A., Kumar, A., Held, I. M., and Gudgel, R.: Statistical–dynamical predictions of seasonal North Atlantic hurricane activity, Monthly Weather Review, 139, 1070–1082, https://doi.org/10.1175/2010MWR3499.1, 2011.

Villarini, G., Luitel, B., Vecchi, G. A., and Ghosh, J.: Multi-model ensemble forecasting of North Atlantic tropical cyclone activity, Climate dynamics, 53, 7461–7477, 2019.

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., and Bretherton, C. S.: Correcting weather and climate models by machine learning nudged historical simulations, Geophysical Research Letters, 48, https://doi.org/10.1029/2021GL092555, 2021.

Wilby, R., Abrahart, R., and Dawson, C.: Detection of conceptual model rainfall—runoff processes inside an artificial neural network, Hydrological Sciences Journal, 48, 163–181, 2003.

Wilby, R. L., Wedgbrow, C. S., and Fox, H. R.: Seasonal predictability of the summer hydrometeorology of the River Thames, UK, Journal of Hydrology, 295, 1–16, 2004.

Wood, A. and Schaake, J.: Correcting Errors in Streamflow Forecast Ensemble Mean and Spread, Journal of Hydrometeorology, 9, 132 – 148, https://doi.org/10.1175/2007JHM862.1, 2008.

World Meteorological Organization: Guidelines on Seasonal Hydrological Prediction (WMO-No. 1274), Tech. rep., 2021.

Wu, Z., Yin, H., He, H., and Li, Y.: Dynamic-LSTM hybrid models to improve seasonal drought predictions over China, Journal of Hydrology, 615, 128 706, 2022.

Xu, T., Longyang, Q., Tyson, C., Zeng, R., and Neilson, B. T.: Hybrid Physically Based and Deep Learning Modeling of a Snow Dominated, Mountainous, Karst Watershed, Water Resources Research, 58, e2021WR030 993, https://doi.org/https://doi.org/10.1029/2021WR030993, e2021WR030993 2021WR030993, 2022.

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montani, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehrhan, J., and Wunram, C.: MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems, Atmospheric Science Letters, 9, 80–87, https://doi.org/https://doi.org/10.1002/asl.183, 2008.

Zhang, B., Wang, S., Qing, Y., Zhu, J., Wang, D., and Liu, J.: A vine copula-based polynomial chaos framework for improving multi-model hydroclimatic projections at a multi-decadal convection-permitting scale, Water Resources Research, p. e2022WR031954, 2022.

Zhang, W., Villarini, G., Slater, L., Vecchi, G. A., and Bradley, A. A.: Improved ENSO forecasting using bayesian updating and the North American multimodel ensemble (NMME), Journal of Climate, 30, 9007–9025, 2017.

Acknowledgements. We wish to thank two anonymous reviewers for their insightful comments that helped improved the work. LJS acknowledges support from UK Research and Innovation (MR/V022008/1 and NE/S015728/1). AYYC and MZ acknowledge support from the WSL MaLeFiX project within the program 'Extremes'. LA is supported by the Canada First Research Excellence Fund's Global Water Futures programme. GV acknowledges support from the USACE Institute for Water Resources. CM acknowledges support from Science Foundation Ireland (SFI/17/CDA/4783).