Title
Can machine learning accelerate process understanding and decision-relevant predictions of river water quality?

Journal
Hydrological Processes, 36(4)

ISSN
0885-6087

Authors
Varadharajan, Charuleka
Appling, Alison P
Arora, Bhavna
et al.

Publication Date
2022-04-01

DOI
10.1002/hyp.14565

Peer reviewed

Received: 1 September 2021 | Revised: 17 January 2022 | Accepted: 19 March 2022

DOI: 10.1002/hyp.14565

THE ROLE OF BIG DATA AND MACHINE LEARNING IN HYDROLOGICAL PROCESSES

WILEY

Can machine learning accelerate process understanding and decision-relevant predictions of river water quality?

Charuleka Varadharajan[1] | Alison P. Appling[2] | Bhavna Arora[1] | Danielle S. Christianson[3] | Valerie C. Hendrix[3] | Vipin Kumar[4] | Aranildo R. Lima[5] | Juliane Müller[3] | Samantha Oliver[2] | Mohammed Ombadi[1] | Talita Perciano[3] | Jeffrey M Sadler[2] | Helen Weierbach[1] | Jared D. Willard[4] | Zexuan Xu[1] | Jacob Zwart[2]

[1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California, 94010, USA

[2]U.S. Geological Survey, Water Mission Area, Reston, Virginia, USA

[3]Computing Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA

[4]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota, USA

[5]Aquatic Informatics, Vancouver, BC, V6E 4M3, Canada

Correspondence
Charuleka Varadharajan, 1 Cyclotron Road, Berkeley, CA-94720, USA.
Email: cvaradharajan@lbl.gov

Funding information
U.S. Department of Energy, Office of Science, Biological and Environmental Research, Grant/Award Number: DE-AC02-05CH11231; National Science Foundation, Directorate for Computer and Information Science and Engineering, Grant/Award Number: 1934721; Lawrence Berkeley National Laboratory; U.S. Geological Survey

Abstract

The global decline of water quality in rivers and streams has resulted in a pressing need to design new watershed management strategies. Water quality can be affected by multiple stressors including population growth, land use change, global warming, and extreme events, with repercussions on human and ecosystem health. A scientific understanding of factors affecting riverine water quality and predictions at local to regional scales, and at sub-daily to decadal timescales are needed for optimal management of watersheds and river basins. Here, we discuss how machine learning (ML) can enable development of more accurate, computationally tractable, and scalable models for analysis and predictions of river water quality. We review relevant state-of-the art applications of ML for water quality models and discuss opportunities to improve the use of ML with emerging computational and mathematical methods for model selection, hyperparameter optimization, incorporating process knowledge into ML models, improving explainablity, uncertainty quantification, and model-data integration. We then present considerations for using ML to address water quality problems given their scale and complexity, available data and computational resources, and stakeholder needs. When combined with decades of process understanding, interdisciplinary advances in knowledge-guided ML, information theory, data integration, and analytics can help address fundamental science questions and enable decision-relevant predictions of riverine water quality.

KEYWORDS
analysis, machine learning, models, predictions, rivers, streams, water quality

1 | INTRODUCTION

Water quality refers to physical, chemical, and biological characteristics of water bodies. Water quality is an important aspect of our ability to use rivers and streams for drinking water, healthy aquatic ecosystems, farmland irrigation, and other beneficial purposes. River water quality has been declining globally over the past century due to population growth and increasing urbanization, industrialization, and

agriculture (Abbott et al., 2019). These stressors have led to salinization and alkalinization of rivers (Kaushal et al., 2013, 2021), nutrient runoff causing algal blooms and eutrophication (Hartmann et al., 2014; McDowell et al., 2020), and river contamination from persistent organic compounds, pesticides, and toxic metals (Schwarzenbach et al., 2010). Climate change and extreme events such as floods, droughts, and wildfires are projected to exacerbate problems of water quality by decreasing flows, changing biogeochemical cycles, and increasing contaminant concentrations (Lyubimova et al., 2016; Murdoch et al., 2000; Nilsson & Renofalt, 2008; Whitehead et al., 2009). The deterioration of river water quality has direct consequences for aquatic and human health. The costs of watershed management to comply with water quality regulatory criteria are significant, even when the benefits are uncertain (Keiser et al., 2019). Thus, water resource managers rely on monitoring and simulations of hydrological processes to make optimal decisions, and are helped by models that account for complex biogeochemical processes and their climactic, hydrological and human drivers.

Different classes of water quality models can be used for predictions depending on the relevant water quality variables, spatial scales (e.g., reach, watershed, regional), and complexity (Rode et al., 2010), some of which are described in recent reviews (Burigato Costa et al., 2019; Fu et al., 2020). High-fidelity mechanistic models can incorporate the tight coupling of physical and biogeochemical processes such as climatic variations, hydrologic fluxes, weathering, and biological interactions on solute transport and mixing at pore to subcatchment scales. For example, integrated surface-subsurface hydrologic models (e.g., Amanzi/ATS) coupled with reactive transport models (RTMs) can attain spatially explicit representations of critical zone components, and provide accurate estimates of geochemical exports to rivers at reach to hillslope scales (Arora, Spycher, et al., 2016; Dwivedi et al., 2018; Steefel et al., 2015; Xu et al., 2021). Despite the demonstrated success of these models, they face important challenges at larger spatial domains in part due to the computational expense from having highly resolved spatial grids and complex biogeochemical processes needed to represent heterogeneous watershed characteristics influencing water quality (Steefel, 2019). Capabilities for upscaling current RTMs and representation of geochemistry in larger scales are in their early stages of development (Arora et al., 2015; Jan et al., 2021; Li et al., 2021).

At watershed- to basin-scales, models such as the semi-empirical soil and water assessment tool (SWAT; Arnold et al., 1998) and hydrologic simulation program Fortran (e.g., HSPF; Bicknell et al., 1996) have been widely used for surface water quality modelling (Fu et al., 2019). The models contain elements called hydrologic response units (HRUs) that are non-continuous areas within a sub-basin grouped by similar properties such as land use, soil, or topography based on user-defined thresholds (Chen, Xu, et al., 2019; Her et al., 2015). This discretization approach makes the models computationally efficient but introduces limitations such as the lack of hydrologic connectivity between the HRUs and sensitivity of the models to the HRU definitions, although some of these limitations are addressed in newer versions of the code (e.g., SWAT+; Bieger et al., 2017; Her et al., 2015; Paul et al., 2019). The models also require watershed-specific calibration and validation, and the parameters can have substantial uncertainties due to equifinal solutions (Van Liew et al., 2003; White & Chaubey, 2007; Xie & Lian, 2013).

Most regional- to continental- scale process-based models that do capture geochemistry (e.g., WITCH; Roelandt et al., 2010) are zero-dimensional box-models, which are not adequate for representing complex flow and chemical transport dynamics (Li et al., 2017). Process-based models have been used to predict physical variables such as stream temperature (e.g., MOSART-heat, PRMS-SNTemp, VIC-RBM; Li et al., 2015; Sanders et al., 2017; van Vliet et al., 2012) at these spatial scales by incorporating thermodynamics and energy balance into terrestrial hydrology models. Alternatively, statistical models are also used for regional- to continental-scale predictions of water quality, although these have other significant limitations. For example, the SPARROW model (Schwarz et al., 2006) has been widely used for spatial water quality predictions at regional scales in the United States (U.S.) but assumes long-term steady-state behaviour, and efforts to incorporate temporal dynamics are nascent (Chanat & Yang, 2018). Multi-linear regression and auto-regressive integrated moving averages (ARIMA) are commonly used methods for temporal predictions of water quality variables but have limited applicability in complex systems with non-linear and non-stationary processes (Chen et al., 2020).

Apart from predictions, numerous studies have used statistical regression to analyse water quality trends and attribute drivers at local to continental scales (e.g., Guo et al., 2019; Kaushal et al., 2018; Monteith et al., 2007; Murphy, 2020; Murphy & Sprague, 2019). Generalized additive models that extend linear models are another statistical approach used to determine water quality trends (e.g., Morton & Henderson, 2008; Yang & Moyer, 2020) and can be adapted to large datasets (Wood et al., 2015). Another common statistical technique is correlation analysis between riverine chemical concentrations and discharge (typically referred to as C-Q relationships), which has been used for a variety of purposes such as examining constituent dynamics at short and long time-scales (e.g., Arora et al., 2020; Evans & Davies, 1998; Godsey et al., 2009; Moatar et al., 2017; Musolff et al., 2021), identifying sources and pathways of different solutes (see Musolff et al., 2021 and references therein), and analysing water quality monitoring data for watershed management (Bieroza et al., 2018; Pohle et al., 2021; Westphal et al., 2020). However, complex C-Q patterns, such as those resulting from variable lags between the hydrograph and chemograph, are difficult to interpret and attempts to link insights gained from C-Q analysis to models have been limited (Liu, Birgand, et al., 2021). Other statistical approaches such as Bayesian hierarchical models (Guo et al., 2019; Rode et al., 2010), wavelets (Arora, Dwivedi, et al., 2016; Parmar & Bhardwaj, 2013), and principal component analysis and non-negative matrix factorization (Liu, Zhang, et al., 2021; Shaughnessy et al., 2021) are used to identify key catchment attributes, sources, and ecosystem control points (Bernhardt et al., 2017) that influence stream water chemistry.

Despite the variety of statistical and modelling approaches available, it is still challenging to make accurate and timely water quality predictions, particularly at large spatial scales, due to the variability of

water quality with space, time, and disturbance (natural or human), and the effect of local characteristics and complex processes on solute transport. There is consensus that water quality modelling has plateaued, with recent studies focusing on incremental improvements to existing models and case study demonstrations for individual basins rather than transformational advances that address the challenges involved (Fu et al., 2020). The field is fragmented, with the choice of model varying substantially depending on the spatial scale and parameters being modelled, as well as access to computation and modelling expertise (Fu et al., 2020), limiting our ability to make decision-relevant predictions.

Recent advances in machine learning (ML) and artificial intelligence (AI) spurred by the increasing availability of public datasets, development of software frameworks, and cloud computing resources have the potential to stimulate a new class of water quality models that can be run at decision-relevant scales, resolutions, and lead times. Artificial intelligence encompasses technologies that can understand inputs from the environment, reason and take actions to meet a performance objective, while ML (the focus of this paper) is a subset of AI that focuses on learning patterns from complex data (Russell & Norvig, 2020). Thus ML is useful for problems that require predictions, deriving insights from data, and decision making (Jordan & Mitchell, 2015). ML applications for geosciences broadly span modelling, automation tasks and data-driven discovery, and include approaches such as classification, spatial or temporal regression, and anomaly detection (Bergen et al., 2019; Reichstein et al., 2019). ML methods include supervised learning, wherein known observations are used to train a model to learn patterns or make predictions on a target variable; unsupervised learning, wherein the algorithm learns patterns or structures in the data heuristically without apriori knowledge of the target; and reinforcement learning, wherein the model learns to make decisions under uncertainty through trial and error interactions with its environment to maximize a reward (see Shen, 2018; Bergen et al., 2019; Xu & Liang, 2021 and references therein for an overview of ML in hydrology and geosciences). Traditional ML methods (also referred to as classical or conventional ML) include support vector machines (SVM), decision trees such as random forests (RF) and gradient boosted trees, and simple artificial neural networks (ANN; e.g., feed forward ANN, extreme learning machines, multi-layer perceptron). Deep learning (DL) models are neural networks (NN) with multiple processing layers that make it possible to learn complex patterns from large datasets and have high predictive skill. Common DL architectures include the convolutional neural network (CNN), recurrent neural networks (RNN) of which the long short term memory (LSTM) network is a popular choice, deep belief networks (DBN) that are composed of restricted Boltzmann machines (RBM), and sparse autoencoders (LeCun et al., 2015; Shen, 2018; Shrestha & Mahmood, 2019). In this paper, we use the term "machine learning" to refer to both traditional ML and DL models.

Over the past decade, there has been a large increase in the use of ML for hydrological predictions (Xu & Liang, 2021) to the point where its adoption in future physical hydrology models seems inevitable (Nearing, Kratzert, et al., 2021). A wide variety of ML approaches have been explored for water quality modelling in rivers across the world, and collectively indicate their potential to extract scientific knowledge and enable optimal management of water quality (Section 2). Yet the rapidly increasing use of ML raises several questions. First, given the diversity of modelling approaches, what are the situations where ML can add value and what are its limitations? What are the primary considerations for model and feature selection? How do we compare performance and evaluate whether models have been constructed and parameterized appropriately? How can ML approaches complement process-based models to improve our prediction capabilities? And most importantly, how can we trust and use models that make predictions based on machine-derived information that is beyond human comprehension?

In this paper, we discuss the potential for using ML in water quality modelling for decision-relevant predictions. We first provide a brief review of ML methods used for several water quality applications that demonstrate its value for modelling and knowledge discovery (Section 2). We then present opportunities for advancement of ML model architectures, transferability, and interpretability, which includes the integration of process knowledge into model design (Section 3). We finally present considerations for the choice of model based on the scale and complexity of the problem being addressed, and availability of data and computational power (Section 4). We posit that the time is ripe for *judiciously* incorporating ML into riverine water quality models for the purposes of improving predictive capabilities and deciphering the complex, diverse human-natural processes that affect water quality. Throughout, we use the term "models" to refer to codes that both generate predictions and analyse data to extract information.

2 | STATE-OF-THE-ART MACHINE LEARNING IN RIVER WATER QUALITY MODELS

Machine learning models have been used for predictions of many water quality variables using various approaches from simple models to hybrid methods combining ML with process-based models (Table 1). Stream temperature is a widely measured physical water quality variable and has been successfully predicted using different ML methods. For example, classical ML methods have been used for monthly to daily predictions of stream temperatures in catchments with different characteristics, and include support vector regression (SVR; Rehana, 2019; Weierbach et al., 2022); decision tree-based regression models such as RF, XGBoost, and their variations (Feigl et al., 2021; Lu & Ma, 2020; Weierbach et al., 2022); and simple ANNs (Feigl et al., 2021; Zhu & Piotrowski, 2020). Among DL models, the LSTM deep neural network has become an increasingly popular choice for regional-scale hydrological predictions due to its ability to encode prior system states in the cell memory (e.g., Kratzert et al., 2018). For example, Rahmani, Lawson, et al. (2021) originally used an LSTM network to predict daily stream temperatures for 118 pristine catchments in the continental U.S. and in a subsequent study extended the LSTM approach to make predictions in data-sparse, unmonitored and dammed catchments (Rahmani, Shen, et al., 2021). Approaches that

TABLE 1 Examples of water quality applications where machine learning has been applied

Application	Stream water quality variables	Classical ML (single or ensemble)	Deep learning	Hybrid process + ML	Unsupervised learning
Time-series predictions	Temperature, DO, conductivity, nutrients, pH, turbidity, cations, anions, dissolved organic matter, BOD, COD, chlorophyll-a	SVR; decision trees, ANN, GP, fuzzy logic, ensembles (Chen et al., 2020; Feigl et al., 2021; Khullar & Singh, 2020; Rajaee et al., 2020; Rehana, 2019; Weierbach et al., 2022; Zhu & Piotrowski, 2020)	LSTM (Rahmani, Lawson, et al., 2021; Zhi et al., 2021); LSTM+CNN (Baek et al., 2020), DBN (Solanki et al., 2015; Yan et al., 2020)	LSTM and graph RNN combined with stream temperature process model (Jia, Zwart, et al., 2021); ANN + salinity process model (Hunter et al., 2018)	
Predictions in ungaged basins	Temperature	XGBoost, SVR (Weierbach et al., 2022)	LSTM (Rahmani, Shen, et al., 2021)		
Short-term predictions, surrogates for monitoring and detection	Temperature, DO, pH, conductivity, turbidity, nutrients, water quality indices	SVM, RF, ANN, ensembles (Bui et al., 2020; Green et al., 2021; Harrison et al., 2021; Lu & Ma, 2020; Paepae et al., 2021)	LSTM (Liu et al., 2019)		Dynamic time warping clustering (Lee et al., 2020)
Event dynamics and classification	Suspended sediments	Restricted Boltzman machines, Bayesian belief networks (Hamshaw et al., 2018, 2019; Murray et al., 2012)	CNN (Hamshaw et al., 2019)		Multivariate event time series clustering (Javed et al., 2021)
Process understanding and knowledge discovery	Temperature, pH, TDS, conductivity cations, anions, nutrients, *Escherichia coli*	Decision trees, NN (Álvarez-Cabria et al., 2016; Mirzaei et al., 2020; Povak et al., 2014; Wang et al., 2021)			
Remote sensing estimates	Suspended and other sediments, turbidity, DO, metals, nutrients, chlorophyll-a, organic matter, blue-green algae	ANN, SVM, GP, decision trees (Topp et al., 2020)	Progressively deep NN (Peterson et al., 2020)		
Decision support and management scenarios	Conductivity, pH, DO, nutrients, suspended sediments, anions, chlorophyll	Bayesian networks/ probabilistic graph models (Forio et al., 2015; Mount & Stott, 2008; Phan et al., 2019)			

Note: See references for further details on techniques used for specific water quality variables.

combine process-based models with DL have also been used for basin-scale stream temperature predictions, particularly in regions with sparse data (e.g., Jia, Zwart et al., 2021, Section 3.2). In all cases, the ML approaches were able to improve prediction accuracies compared to base statistical and process-based models.

ML models have also been used for predictions of other water quality variables and indices such as dissolved oxygen (DO), electrical conductivity, total dissolved solids (TDS), pH, nutrients, suspended sediments and turbidity, cations, anions, biological oxygen demand (BOD), chemical oxygen demand (COD), and chlorophyll. Recent

reviews identify popular classical ML methods for time-series predictions including ANN, decision trees, SVM, genetic programming (GP), fuzzy logic, and hybrid approaches such as genetic algorithm-NN, adaptive neuro-fuzzy inference systems, and wavelet-NN (Chen et al., 2020; Khullar & Singh, 2020; Rajaee et al., 2020). Hybrid models with ensembles of standalone ML models using methods such as bagging and random subspace (Melesse et al., 2020) or combining ML with process-based models (Hunter et al., 2018) have been used to predict salinity at river basin- to reach-scales in Iran and Australia. Examples of recent efforts to incorporate DL approaches for these predictions include the use of an LSTM to predict DO concentrations in 506 pristine U.S. catchments achieving moderate accuracy (Zhi et al., 2021), and an LSTM paired with a CNN (that generated streamflow estimates) to predict total nitrogen, phosphorus, and organic carbon in a major Korean river basin (Baek et al., 2020). LSTM, RF, and hybrid ML models have also been used for short-term predictions of a suite of water quality variables (e.g., DO, pH, conductivity, turbidity, nutrients, water quality indices) with high-frequency monitoring; in some cases classical ML surrogate models are used as soft sensors to make predictions of variables that are difficult or laborious to measure directly such as those that require laboratory sample analysis (Bui et al., 2020; Green et al., 2021; Harrison et al., 2021; Liu et al., 2019; Lu & Ma, 2020; Paepae et al., 2021).

In addition to predictions, ML approaches are used to discover understanding about the dynamics of water quality and underlying sources, drivers, and mechanisms. A common use of ML is to detect events of interest from high-frequency data obtained from sensor networks. For example, insights into episodic sediment dynamics from hundreds of storm events were derived using RBMs and CNNs to automate analysis of C-Q plots (Hamshaw et al., 2018, 2019). Classification and clustering pattern recognition techniques have been used to identify similarities between monitoring stations (Lee et al., 2020), hydrological events (Javed et al., 2021), and contaminant sources (Vesselinov et al., 2018). Classical ML methods such as decision tress, NN, and self-organizing maps have been used to identify landscape attributes and human factors that affect stream water quality (Álvarez-Cabria et al., 2016; Mirzaei et al., 2020; Wang et al., 2021), and characterize relevant processes such as cation weathering rates (Povak et al., 2014).

Another application area is the use of ML for analysis and estimation of inland water quality parameters obtained from remote sensing imagery. The estimates are most commonly derived using classical ML approaches, but are limited to larger rivers and a few variables due to the spatial resolution of land surface sensors (~10–30 m), and broad spectral bands (Hassan & Woo, 2021; Hestir et al., 2015; Topp et al., 2020). Peterson et al. (2020) demonstrated that a progressively deep NN model outperformed SVR, extreme learning machine, and multi-linear regression in estimating water quality parameters such as blue-green algae, chlorophyll-a, fluorescent dissolved organic matter (FDOM), DO, conductance, and turbidity in the Upper Mississippi River using data from Landsat-8 and Sentinel-2. A harmful algal bloom (HAB) detection system HABNet (Hill et al., 2020) for coastal waters used a CNN combined with an LSTM, SVM, or RF applied on data from moderate resolution imaging spectroradiometer (MODIS) data. Remote sensing datasets relevant to river water quality are likely to grow in the future with newer high-resolution satellites, instruments with hyperspectral bands, and increasing use of unmanned autonomous vehicles (UAV) to acquire water quality data (Sibanda et al., 2021; Topp et al., 2020), and can potentially address data gaps for water quality ML (Section 4.4).

Finally, probabilistic graph models such as Bayesian networks have been used extensively for decision support in water management (Phan et al., 2019). Bayesian networks can integrate both quantitative data (e.g., time-series of measured concentrations and discharge), and qualitative data (e.g., expert knowledge and stakeholder beliefs), and can generate probabilistic predictions for different management and climate scenarios. For example, Bayesian networks were used to design rules for trading transferable pollutant discharge permits in rivers (Mesbah et al., 2009), for real-time monitoring and contaminant warning systems (Murray et al., 2012), and to predict water quality in different types of catchments (Forio et al., 2015; Mount & Stott, 2008).

These examples illustrate the value of using ML for a variety of global river water quality modelling and monitoring applications. The use of DL and hybrid models combining ML and process-based codes is still nascent and is expected to grow in the near future. In the next two sections, we describe how the appropriate use of ML for different water quality applications can best incorporate recent computational and mathematical developments due to the large number of variables of interest, heterogeneity of human and natural processes that influence water quality, and scarcity of available data.

3 | OPPORTUNITIES FOR ADVANCEMENT OF WATER QUALITY ML

ML methods have advanced rapidly over the past decade, and diverse approaches can be used to improve model performance, efficiency, and robustness. Here we highlight particularly promising techniques, many of which have been used in hydrological applications, to address questions raised in Section 1 related to the use of ML for water quality modelling (also see Varadharajan et al., 2021).

3.1 | How do we select and design optimal model architectures?

ML model selection and architectures (e.g., number of layers, batch size, nodes for neural networks) can affect model performance and computational costs. The choice of the best performing model (ML, statistical, or process-based) will depend on the metrics considered including model accuracy, variance, robustness, bias, and computational speed (Belitz & Stackelberg, 2021; Kratzert, Klotz, Shalev, et al., 2019). A common approach for model selection in many of the studies cited above (Section 2) is to train different model types, and choose the model with the greatest predictive skill. Another approach

to model selection is to *not* choose, i.e., to create model ensembles that exploit the diversity of predictive skill from different models. Model ensemble approaches are common in hydrology (Fleming et al., 2021). Specifically, ML ensembles have shown promise in improving predictions (e.g., flood forecasting), quantifying uncertainty (Fleming et al., 2015; Jiang et al., 2014; Zounemat-Kermani et al., 2021), and are now being incorporated into operational models (Fleming et al., 2015, 2021). Ensembles can be formulated in many ways, from traditional methods such as bagging, boosting, dagging, model averaging, and stacking to more sophisticated combinations such as running simulations on different classes of models or pre-training ML models on different process models (e.g., Section 4.2). Ensemble combinations of multiple models can be even more beneficial when the errors of individuals models are not correlated (Hsieh, 2009).

Besides model selection, other important choices in the application of ML are selection of optimal input features, model training, and architecture design. In regression and classification, a best practice is to use k-fold cross-validation that partitions available data into k-sets and iteratively trains the model using data from each set for testing, which leads to better model generalizability (Bergmeir & Benítez, 2012). Common pitfalls in model design are including excess irrelevant or redundant variables as inputs, variable selection bias (i.e., using the same data for training and inputs), resubstitution validation (i.e., testing the model with training data), use of inconsistent cross-validation and resampling procedures across model architectures being implemented, and data leakage (e.g., using testing data for model training or hyperparameter optimization, or pre-processing the entire dataset prior to splitting the data into cross-validation folds), which can lead to overfitting (Gharib & Davies, 2021; Zhang, 2007). These pitfalls can be avoided by understanding the details and limitations of the models being implemented, following best practices, and using robust ML workflows (Gharib & Davies, 2021; Zhang, 2007).

Different hyperparameter optimization (HPO) methods have been developed to find the best-performing ML model architectures (Yu & Zhu, 2020). In many applications such as those cited above, the hyperparameters are either not chosen (i.e., default hyperparameters selected), hand-tuned (trial and error of different architectures), or selected using a grid or random search approach (Bergstra et al., 2011). In a few models (e.g., RF), default hyperparameters can work well (Probst et al., 2019); however, in general HPO is expected to improve model performance. Although random search has been shown to outperform grid search (Bergstra & Bengio, 2012), both approaches require a time-consuming approach of training each architecture, which can be computationally expensive, result in inferior performance, and limit the number of architectures that can realistically be explored. Instead, efficient solution methods using Bayesian optimization methods such as the Tree-of-Parzen-Estimators in the Python package Hyperopt can be applied (Bergstra et al., 2013). Müller et al. (2020) demonstrate another derivative-free Bayesian optimization method using surrogate models, such as Gaussian process or radial basis functions, to map the architecture search space to its performance and allows efficient exploration of the space by adaptive sampling approaches. Near-optimal hyperparameters for different NN including DL architectures can be found with relatively few evaluations using this approach (e.g., Figure 1).

Automated machine learning (AutoML) frameworks that use concepts of ensembles and automatic HPO are emerging as a promising approach for reducing the barrier to adoption of ML, and have shown improvements in the efficiency and robustness of ML algorithms (Feurer et al., 2015; Hutter et al., 2019). A forward-looking automated ML approach for model design is "neural architecture search" (Ren et al., 2021) where algorithms automatically design model architecture. This approach has been successfully applied to algal classification using CNNs (Park et al., 2019) and spatiotemporal predictions (Li et al., 2020), and is even being considered for physics-based learning models (Ba et al., 2019).

3.2 | How do we incorporate scientific knowledge into ML models?

Traditional ML models can produce physically inconsistent results (Karpatne et al., 2017; Kashinath et al., 2021) as they only look for statistical relationships in the training data and are unable to extrapolate outside of the training dataset used for building the ML model. The sparseness of water quality and related environmental datasets, and projected changes in climate and land use make it challenging to use ML models per se for predictions in unmonitored regions or for long-term projections because they can only be trained, validated, and evaluated on past data (Duan et al., 2020; Kratzert, Klotz, Brandstetter, et al., 2019; Xu & Liang, 2021).

The scientific community has been converging on the use of knowledge-guided ML (KGML; see Figure 2 as an example) to make physically consistent and generalizable predictions (including out-of-bounds estimates) and decrease training times (Kashinath et al., 2021). KGML approaches have been used to incorporate scientific knowledge into ML model training and predictions in several ways (Willard et al., 2022) that include (i) modification of loss functions to penalize the model when violating known physical laws and relationships (Jia et al., 2019; Karpatne et al., 2018), (ii) training or pre-training ML models on process-based model output (Konapala et al., 2020; Read et al., 2017), (iii) incorporating differential equations in ML models (Chen, Rubanova, et al., 2019; He et al., 2020; Rackauckas et al., 2020), (iv) multi-task learning for multi-objective optimization (Sadler et al., 2022), (v) modifying ML model architecture to impose symmetry or better represent the system being modelled (Daw et al., 2020; Khandelwal et al., 2020; Kunin et al., 2021; Ling et al., 2016), and (vi) most commonly surrogate (or reduced-order) models that emulate process model behaviour with considerably less computational expense (Xu & Liang, 2021). Hybrid KGML methods have also been used for hypothesis generation and establishing causal relationships between system drivers and responses (Tsai et al., 2020).

Numerous combinations of LSTM models and their variants with different process models have been shown to improve predictive

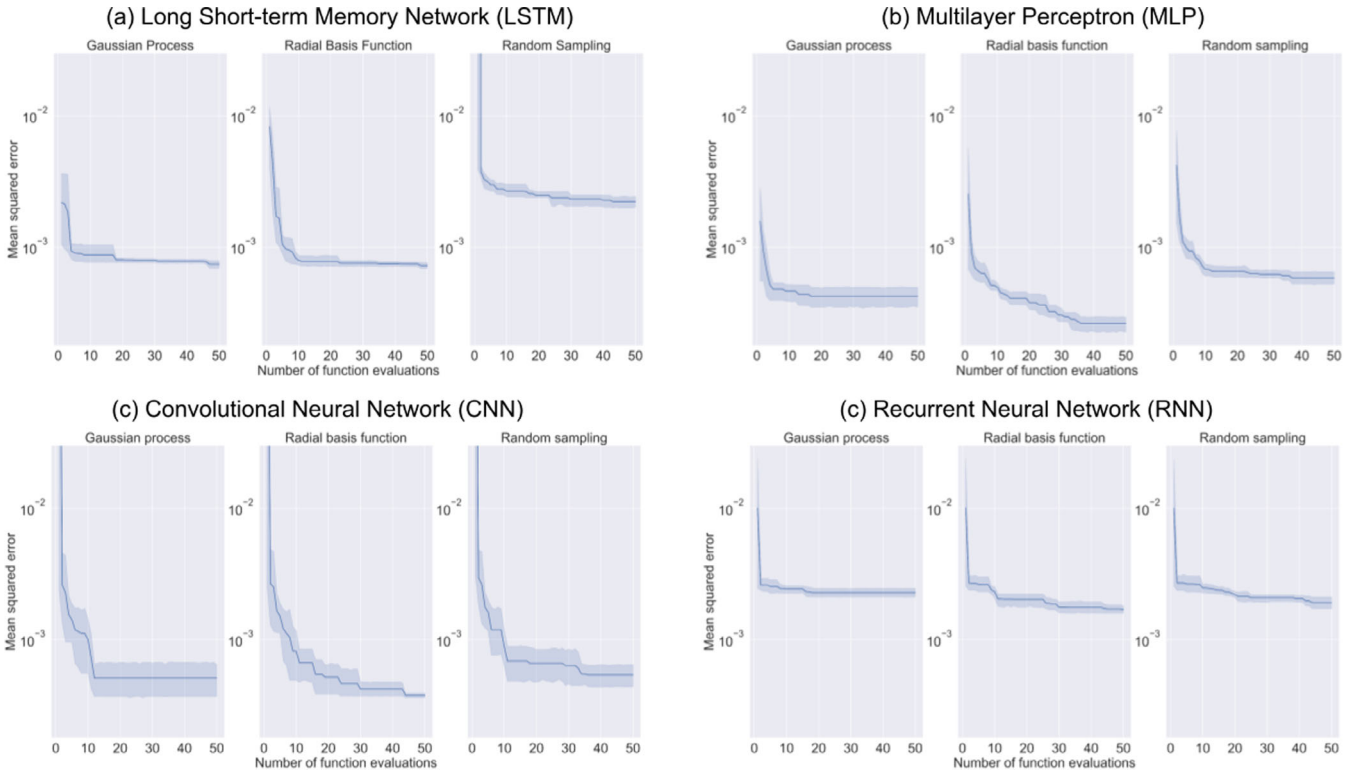

FIGURE 1 Bayesian optimization approaches using surrogate models such as Gaussian process and radial basis functions enable faster convergence to the optimal set of hyperparameters for deep learning models used to predict groundwater levels (figure modified from Müller et al., 2020)

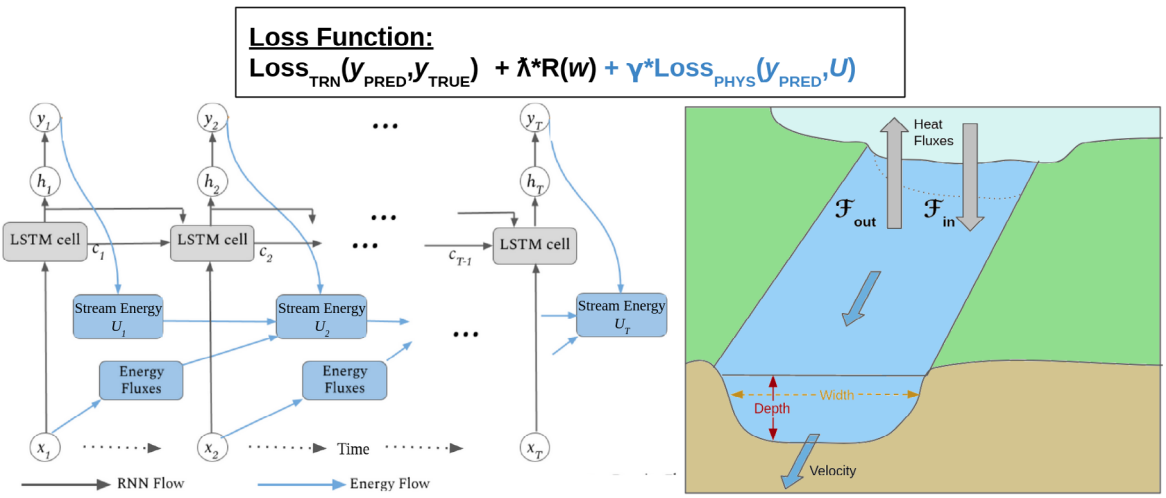

FIGURE 2 Overview of knowledge-guided machine learning (ML) approaches incorporating physical constraints into the loss function of a long-short term memory (LSTM) network used to model stream temperatures. The loss function penalizes the ML model for errors in predicting stream temperature (yPRED = model-predicted stream temperature; yTRUE = observed stream temperature) and stream thermal energy (*U*), the latter of which cannot be created or destroyed from one time step to the next according to thermodynamics. Energy fluxes can be transformed or transported into (fin) or out of (Fout) the stream system according to equations in the process-based model. *R*(*W*) represents a standard complexity regularization on the model parameters weighted by λ, and Loss_{PHY} aims to ensure consistency with physics by balancing the energy fluxes and is weighted by γ. Figure modified from Willard et al. (2022) and https://www.usgs.gov/media/images/diagram-channel-cross-section-subsections

performance for streamflow and river and lake water quality including out-of-bounds predictions in comparison to base process or LSTM models (Frame et al., 2021; Hanson et al., 2020; Jia, Zwart, et al., 2021; Konapala et al., 2020; Read et al., 2017). For example, Jia et al. (2021) used a modified graph-based LSTM that included representations of the river network hydrography, pre-trained with output

from a process-based model PRMS-SNTemp (Sanders et al., 2017) to predict daily stream temperatures across the Delaware River Basin. This study demonstrated that the hybrid model was able to outperform base LSTM models in terms of prediction accuracies and the ability to extrapolate when data were sparse (Figure 3).

Other KGML approaches include variations of GP, a supervised ML technique that mimics biological evolution. The advantages of GP include its ability to produce explicit mathematical relationships between input and output variables, add constraints based on physical realistic possibilities, and generate interpretable results. Genetic programming has been used extensively in hydrology to build interpretable rainfall-runoff, groundwater, and water quality models; estimate evapotranspiration; and extract information about reservoir operations (Danandeh Mehr et al., 2018; Fallah-Mehdipour & Haddad, 2015; Herath et al., 2021). KGML versions of GP models include a tree-adjoining grammer (TAG) formalism for representing dynamic processes that has been demonstrated for predictions of phytoplankton biomass in rivers (Park et al., 2021).

3.3 | How can ML be used to make generalizable predictions with sparse data?

Water quality data tend to be sparse (Section 4.4), and there is a need to transfer models from data-rich, small-scale monitored sites to other regions in a manner that accounts for the spatial heterogeneity of watershed characteristics. Recent studies have demonstrated the use of LSTM and ensemble XGBoost models for predictions of streamflow and stream temperature in unmonitored basins (Kratzert, Klotz, Herrnegger, et al., 2019; Rahmani, Shen, et al., 2021). Another ML approach to generalize models is Transfer Learning (Weiss et al., 2016), which translates models built on highly observed locations to other sites that are data-sparse or unmonitored, although the challenge lies in choosing which models to transfer to a given site. Recently, meta transfer learning (MTL) has been proposed as a means to address this problem, wherein multiple base models (ML or site-specific process model) can be trained or calibrated for different tasks such as prediction at a given location, and then a meta-learning model (de Oliviera, 2019;

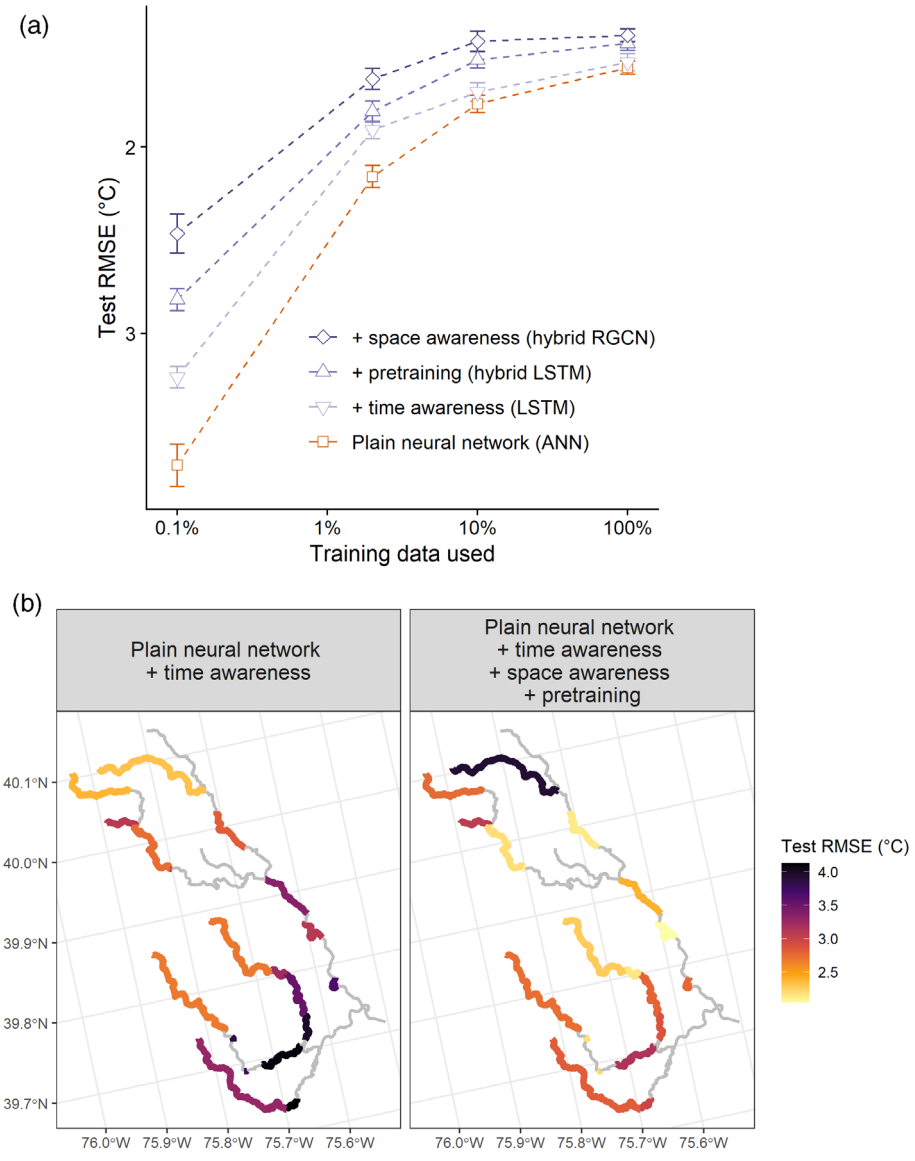

FIGURE 3 (a) Comparison of pure machine learning (ML) approaches using artificial neural networks (ANN) to various process inclusions (time and space awareness, pre-training with a process model) for predicting stream temperature in the Delaware River basin (data from Jia et al., 2021). The error bars in panel a represent the standard deviation of test root mean squared error (RMSE) across five model runs. Panel (b) shows model performance comparison between a neural network with time awareness (i.e., long short-term memory network; long short term memory (LSTM); left) and a hybrid graph neural network (recurrent graph convolutional network; RGCN; right) with time and space awareness in addition to being pre-trained with a process model. Models shown in panel (b) were trained using 0.1% of available data during the training period and only segments with more than five observations during the test period are shown (data from Jia et al., 2021)

Vanschoren, 2019) predicts which base model to transfer based on past performance metrics and location-specific attributes (Willard, Read, et al., 2020). For example, Willard, Read et al. (2020) used MTL for predictions of lake temperature in 305 artificially defined "unmonitored" lakes (targets) by transferring process-based and hybrid KGML models calibrated for 145 data-rich lakes (sources) using past performance metrics of the source models along with features such as lake attributes (e.g., surface area, maximum depth), meteorological statistics, and source lake temperature statistics. MTL flexibly allows the use of any type of model built on well-observed locations making it a noteworthy candidate for increased regional scaling.

3.4 | Can ML methods be adapted to predict extreme values?

Typical ML methods have inbuilt assumptions (e.g., the random variables follow a Gaussian distribution), which can lead to poor estimates when they do not apply for some datasets. In particular, the choice of a loss functions can directly affect the results (Bishop, 2006). For example, a squared loss (L2 norm) is more sensitive to outliers than absolute value loss (L1 norm; Breiman & Friedman, 1985; Hastie et al., 2009; Zhang, 2007). When outliers are equally important as in the case of extreme events, it is not uncommon to utilize the probability distribution within the loss function (Cawley et al., 2007; Hsieh, 2009) or choose alternate loss functions such as the relative entropy loss that computes the distance between two distributions (Qi & Majda, 2020).

Many studies in hydrology have attempted to improve predictions of extremes such as floods (Mosavi et al., 2018) using multiple techniques that include using different criteria for model selection (Coulibaly et al., 2001), conditional density estimation networks (Cannon, 2012), training exclusively on extreme events such as historical high-flow data (Fleming et al., 2015), adjustment of ML prediction bias to improve performance on the tails of the distribution (Belitz & Stackelberg, 2021), and using KGML models, for example by training an ML model on simulation data containing extremes that might not exist in the observation data (Read et al., 2017; Xie et al., 2021). However, in some cases KGML models may perform worse than traditional DL models; for example, Frame et al. (2021) found that an LSTM constrained to conserve mass was not able to predict peak flows as well as the base LSTM, although both models had lower errors than the process-based model used for comparison. Although these examples show the potential for using ML for predictions of extreme values, data scarcity for rare events warrants consideration in model choice and design, and alternate ML methods may need to be used (or modified) to handle extreme values (Cannon, 2010).

3.5 | How do we represent complex, hierarchical data in ML models?

The heterogeneity and multi-scale nature of water quality processes and their drivers requires representation of high-dimensional complex data in ML models, but this can be challenging to do in a computationally tractable manner. One approach that has been used is to modify ML models to embed complex spatial information for time-series predictions. For example, Kratzert et al., (2019) developed an entity aware-LSTM to include information on static catchment attributes into the input gate of an LSTM to improve streamflow predictions across 531 pristine basins. Jia et al. (2021) used a recurrent graph convolution network to include spatial information on river segments into the predictions, and in a subsequent study developed a real-time active learning approach that used spatial and temporal information to select representative samples for optimizing model training (Jia, Lin, et al., 2021). Creating advanced multi-scale graph-based data representations that represent river network structure and carry a diverse set of node and edge attributes can enable embedding multimodal, multi-scale, and multi-temporal information, and correlate structures across different scales and time-steps.

Probabilistic graph models (PGMs), which are data structures for encoding probability distributions, can also be used to include such structural information (Koller & Friedman, 2009). In a PGM, the nodes of the graph represent random variables, and the edges represent dependencies between variables. An optimization process using a PGM involves defining an energy function, which can embed contextual information about the data, i.e., relationships among data points at different resolutions, scales, or time steps (prior knowledge). PGMs can be optimized by targeting the maximization of the joint probability (or the minimization of the energy function) of the graph (Laude et al., 2018), which would result in specific river network predictions. Although classes of PGMs (e.g., Bayesian network models) have been used for water quality applications (Section 2), newer models that integrate PGM with DL such as DBN have been used for water quality predictions (Solanki et al., 2015; Yan et al., 2020).

3.6 | How do we build trustworthy and interpretable ML models?

Domain experts often are hesitant to replace process-based models with ML on the grounds that the explainability, interpretability, and trustworthiness of ML frameworks are questionable even if predictions are more accurate (Rudin, 2019). Model transparency and interpretability is especially important for water management decisions that have important societal implications and that have to consider future unknown scenarios.

Recent advances in explainable AI such as local interpretable model-agnostic explanation (LIME; Ribeiro et al., 2016) or Shapley additive explanations based on occlusion analysis (SHAP; Lundberg & Lee, 2017) can explain individual predictions by many ML models (Samek et al., 2021). Both methods have been successfully implemented for explaining predictions in rainfall-runoff modelling (Althoff et al., 2021; Yang & Chui, 2021), and are applicable for water quality applications as well (Wang et al., 2021). Also in the rainfall-runoff domain, Kratzert et al. (2019) used integrated gradients (Sundararajan et al., 2017) to confirm a theory-consistent influence of precipitation and air temperature on a NN state that correlated with

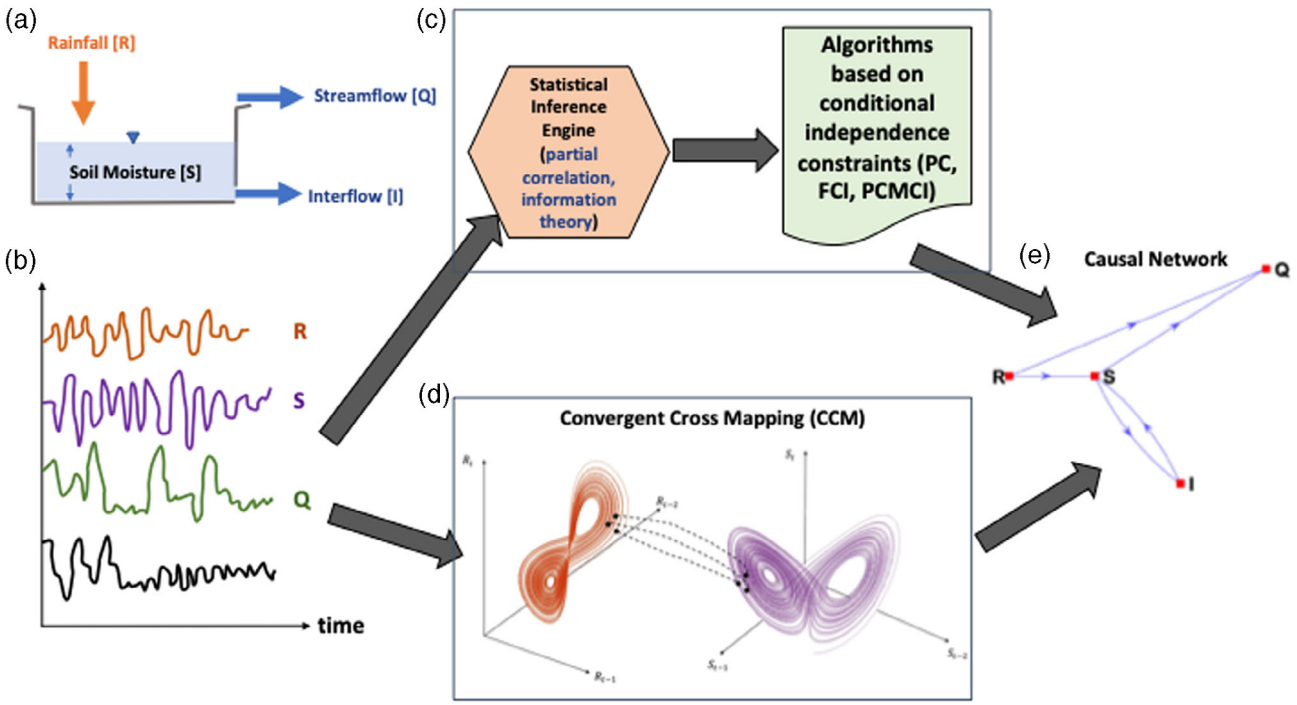

FIGURE 4 Graphical illustration of the use of causal inference methods (highlighted in boxes) for hydrological applications: (a) conceptual understanding of a physical system (i.e., rainfall-runoff process; hydrologic bucket model), (b) observational time series of the four system variables in panel a, (c) a broad class of causal inference algorithms are based on conditional independence constraints such as Granger (Granger, 1969), transfer entropy (Schreiber, 2000), PC algorithm (Spirtes & Glymour, 1991), fast causal inference; FCI (Spirtes & Glymour, 2000), and PCMCI (Runge et al., 2019). Conditional independence is performed using a statistical inference engine (e.g., partial correlation or information theory), (d) an alternative paradigm of causal inference are methods based on time-delay embedding whereby pairwise-bivariate timeseries (e.g., R and S) are used to construct attractors from which causal relations can be assessed (Sugihara et al., 2012), (e) the causal network obtained from causal inference algorithms where nodes represent variables and edges represent causal relations. The obtained causal inference network can be compared and contrasted to the conceptual a priori understanding of the system shown in panel a

snow water equivalent, building trust in the ability of such models to capture known physical processes. More recent work on the same subject used probes (e.g., linear regression models or stacked multilayer perceptrons) to connect LSTM cell states to the output for tracking the evolution of the LSTM during the training process and verifying whether the LSTM learned physically realistic mappings from inputs to outputs (Lees et al., 2021).

Causal inference is another approach to extract information on complex water quality processes and infer the effect of one variable on other system responses (Pearl, 2009; Sugihara et al., 2012). Causal inference methods can be grouped into three categories: (1) algorithms based on conditional independence constraints such as Granger causality (Granger, 1969), transfer entropy (Schreiber, 2000), PC algorithm (Spirtes & Glymour, 1991), fast causal inference (FCI; Spirtes et al., 2001), PC mutual condition information (PCMCI; Runge et al., 2019), (2) methods based on time-delay embedding and chaos theories such as convergent cross mapping (CCM) (Hannart et al., 2016; Sugihara et al., 2012), and (3) approaches based on counterfactual causal theory and structural causal models (Hoyer et al., 2009; Pearl, 2009; Peters et al., 2017). These methods have been adopted in climate and hydrologic sciences (Arora et al., 2019; Hannart et al., 2016; Ombadi et al., 2020; Runge et al., 2019) to extract relationships among system variables and disentangle governing parameters to inform process-based or ML models (e.g., Figure 4).

3.7 | How do we quantify the uncertainties in model predictions?

An important aspect of modelling is uncertainty quantification (UQ), which in ML models includes uncertainties in the data, selection of input features, and model design (Abdar et al., 2021). A common approach to test the robustness of models to inputs is to perform variable feature selection and calculate model errors resulting from choosing different combinations of inputs. Other UQ approaches include Bayesian methods such as Markov chain Monte Carlo, Bayesian active learning, and variational auto encoders and ensemble methods, which can be computationally expensive (Abdar et al., 2021).

Additional UQ challenges to be addressed are the variability of model performance from tuning hyperparameters due to stochastic optimizers used for training (model uncertainty) and use of noisy measurement data (measurement uncertainty). An approach to address

model uncertainty is to use Monte Carlo dropout in model training, wherein each node has a certain probability of being dropped out (dropout rate) and not contributing to the model (Gal & Ghahramani, 2016). Measurement noise should ideally be propagated through the model, but computing required probability densities analytically is possible only for specific types of activation functions (Gast & Roth, 2018; Loquercio et al., 2020). Uncertainty quantification methods that exploit and extend such approaches to obtain confidence intervals for predictions are needed (Dumont et al., 2021).

3.8 | How do we integrate and assimilate relevant datasets into ML models?

Water quality models rely on having access to high-quality, publicly accessible data, and on software tools that enable data integration and assimilation. Many ML studies use readily available off-the-shelf products such as the catchment attributes and meteorology for large-sample studies (CAMELS; Addor et al., 2017), which may impair their extensibility to practical, realistic applications where such data may not be available. Assimilation of new data into models using methods such as ensemble Kalman filters and autoregression (Brajard et al., 2020; Nearing, Klotz, et al., 2021; Zwart et al., 2021), and the use of integrated datasets tailored for the problem can improve prediction outcomes. Software that synthesize data for on-demand queries such as brokering-based tools (Horsburgh et al., 2016; Varadharajan et al., 2022), and methods to streamline quality control and outlier detection, gap-fill, downscale observations, and determine parameters for process models (Bennett & Nijssen, 2021; Campbell et al., 2013; Hill & Minsker, 2010; Leigh et al., 2019; Mital et al., 2020; Russo et al., 2020) would ideally be integrated into ML workflows in parallel with advances in modelling approaches.

4 | CONSIDERATIONS FOR THE USE OF ML IN WATER QUALITY MODELS

Water quality modelling poses several challenges regardless of whether the approach involves ML, statistical, or process-based codes. While each approach has its strengths and limitations, the ultimate model choice, structure, inputs, and parameters would need to take into account the following considerations.

4.1 | Process complexity

River water quality involves a wide range of physical (e.g., temperature, turbidity, conductance), chemical (inorganic and organic), and biological (e.g., bacteria, phytoplankton) parameters. Each parameter can be affected by many natural and anthropogenic factors including solute transport through runoff, groundwater exchange, instream biogeochemical processes, and land use (Lintern, Webb, Ryu, Liu, Bende-Michl, et al., 2018). For example, water temperature is strongly driven by climate but can also be influenced by snowmelt, groundwater influx, and reservoir and power plant operations (Caissie, 2006). Biogeochemical transformations such as redox reactions, uptake of nitrogen, or respiration of dissolved organic carbon add process complexity for non-conservative constituents. Even relatively conserved species such as salts can be affected by processes such as road salt runoff, surface-groundwater interactions, and tidal influence in coastal zones (Kaushal et al., 2018). Disturbances can modify the extent to which different drivers control water quality – e.g., drought increases groundwater influences on salinity and instream biotic influences on nutrient transformation rates (Mosley, 2015). The need for explicit representation of different processes and their feedbacks in mechanistic models has resulted in increasingly complex model structures, and larger parameter spaces that are more difficult to constrain (Rode et al., 2010). ML approaches present an alternative to process models for complex systems where the dynamics are not fully understood, if sufficient data are available through direct or proxy measurements. However, ML and especially DL models with large sets of hyperparameters can also be complex depending on many factors such as model design, optimization process, and data dimensionality (Hu et al., 2021). An important aspect of choosing a model is determining the right level of model complexity required for the decisions to be made. Adding model complexity does not necessarily improve performance, and will increase computational and data requirements (Orth et al., 2015; Rode et al., 2010). Ideally, the choice of a more complex model (e.g., a DL-based architecture) is justified by comparing performance against some baseline lower-complexity ML, statistical, or process models (e.g., multi-linear regression or persistence models).

4.2 | Scale, heterogeneity, and generalizability

The factors influencing riverine water quality can differ in space and time ranging from local-scale surface groundwater fluxes and hyporheic exchange (Sear et al., 1999), reach-scale habitat and riparian vegetation (Newcomer et al., 2021), watershed-scale geomorphology and geology (Burns et al., 2020; Varanka et al., 2015), to regional-scale climate and land use patterns (Guo et al., 2019; Lintern, Webb, Ryu, Liu, Waters, et al., 2018). Building models that translate across spatiotemporal scales and incorporate process understanding of heterogeneous drivers is a significant challenge. Typically models have tried to capture catchment heterogeneity using characterizations of representative field sites from intense, multidisciplinary field investigations (Brantley et al., 2017) or other (e.g., remote sensing) datasets. However, upscaling mechanistic understanding from data-rich small-scale testbeds to other regions has been challenging, in part due to the spatial heterogeneity of watershed characteristics and the need to have computationally tractable models (Tang et al., 2019). Although parameter regionalization and classification of process-based models by catchment have attempted regional scaling with mixed results (Archfield et al., 2015), new approaches to bridge scales for key water quality variables using ML are needed. Although ML models are

typically scale-invariant, their implementation for predictions of a particular spatial and temporal scale may require different model architectures and inputs depending on the processes involved at each scale. Recent hydrological studies explored the use of ML across spatial scales (Gauch et al., 2021; Kratzert, Klotz, Herrnegger, et al., 2019), and approaches such as transfer learning (Section 3.3) can also be used to address this challenge. Finally, the effects of human activity or natural disturbances can persist over long timescales (sometimes decades) requiring some models to include the lagged effects of drivers. For systems where memory effects are significant, ML models that use lagged variables or architectures that support prior states (e.g., LSTM or other sequence models) are warranted.

4.3 | Desired accuracy and computational complexity

A key consideration for model choice is the extent of accuracy and robustness needed to meet the stakeholder objectives (Section 4.5). For example, daily reservoir operations have a much lower tolerance for errors in streamflow and temperature predictions in comparison to longer-term planning for climate change (Culhane et al., 1987). There are tradeoffs to be made between model accuracy, complexity, and computational costs. For example, classical ML models can be trained with substantially lower computational expense and smaller datasets (Hu et al., 2021), and are more interpretable (Rudin, 2019). Deep NN can have lower prediction errors, but have many more parameters than classical ML models thus requiring larger datasets and more computational resources to train, and are more difficult to interpret (Hu et al., 2021; Reichstein et al., 2019).

4.4 | Data availability, integration, processing and representation in models

Riverine water quality modelling typically requires the use of highly diverse multi-scale, multimodal data (e.g., time-series data of flow and solute concentrations, geospatial data such as soil layers and hydrography, remote sensing data, or products of land cover and land use) for parameterization and validation. Model data requirements depend on the complexity, desired scale and resolution, and can be specific to the management question at hand. The amount of data available needs to be considered in choosing the ML model, with classical ML approaches needing less data than DL models. Hybrid KGML models (Section 3.2) and MTL (Section 3.3) can also be better suited for situations with sparse data.

A challenge for water quality modelling is that data are typically sparse because many relevant parameters are measured infrequently using laboratory characterization of manually collected samples, and can be biased due to sampling network design, flow conditions, and time of day when the samples were collected (Smith et al., 1997; Zhang et al., 2019). In-situ sensor technologies that collect high-resolution data only span a limited range of physical and chemical variables (Kruse, 2018), although soft sensor surrogate modelling approaches can be used to generate estimates of variables that are not measured (Section 2). While spatially dense remote sensing observations can be difficult to obtain at sufficiently fine resolutions, there are opportunities to bridge spatial scales by integrating limited point measurements from monitoring networks with remote sensing estimates of water quality (Ross et al., 2019; Topp et al., 2020). Autonomous observations from instruments such as aquatic drones guided by ML (Castellini et al., 2019, 2020) is a future direction that could enable data collection at larger scales.

Discovering, integrating, and processing data for models is also challenging. Typically data of interest are spread across a myriad of sources in different formats, and do not have sufficient metadata, quality, or provenance information to support integration (Larsen et al., 2016). Often co-located datasets on water quality and its drivers are either not available or not easily discoverable. Data also may need to be gap-filled and quality checked prior to use in a model. In particular, water bodies with strong human influence are extremely difficult to model and upscale because predictor datasets on activities such as point source discharge, water withdrawals, and reservoir releases are limited or not easily reusable. Thus many observations are underutilized for ML despite large-scale consolidation efforts such as the Water Quality Portal and GLORICH databases (Hartmann et al., 2014; Read et al., 2017), because it is labor-intensive to harmonize and process data (Shaughnessy et al., 2019; Sprague et al., 2017). Recent efforts to make water data more broadly available and usable such as the U.S. Open Water Data Initiative and the California open water data system are essential for effective management and decision making (Blodgett et al., 2015; Cantor et al., 2021; Larsen et al., 2016). Because data preparation is one of the most time-consuming aspects of model development, the development of benchmark datasets following FAIR (Findable, Accessible, Reusable, Interoperable) principles (Wilkinson et al., 2016), and the use of automated tools that make it easier to discover, synthesize, and assimilate data (e.g., Section 3.8) can accelerate adoption of ML approaches.

4.5 | Data-driven decisions considering stakeholder objectives

Data-driven decision making refers to the concept of making decisions based on data analysis or modelling, and the use of ML for making optimal decisions has been proposed for numerous applications such as business intelligence, healthcare, network optimization, and precision agriculture (Chen, Liu, & Peng, 2019; Liakos et al., 2018; Ma et al., 2020; Provost & Fawcett, 2013). Water management is a complex endeavour, and managers are tasked with making decisions at local (e.g., temperature and salinity control in a reach) to regional (e.g., integrated river basin management) spatial scales. The lead times needed for different decisions can be very different (DeFlorio et al., 2021; Sene, 2016) ranging from months to years for long-term planning (e.g., capital investments, climate adaptation), weeks to

months for seasonal water supply management (e.g., drought and snowmelt forecasting), and hours to weeks for tactical operations and emergency response (e.g., daily reservoir operations, spill mitigation). Water quality management can be especially challenging given the range of variables that are typically regulated (e.g., temperature, salts, oxygen, nutrients, metals, organics) and the diffuse nature of their sources and sinks. Models used for water quality management serve exploratory, planning, or regulatory purposes, which have different requirements for acceptable accuracies, uncertainties, and biases (Harmel et al., 2014).

Ultimately, it is important to consider stakeholder needs and available resources in deciding the appropriate complexity, scale, and resolution of the models. For example, many ML models have been developed for operational decisions that have short lead times (e.g., near-term forecasting), and are typically focused on predictions of a single variable although multitask learning approaches are being explored (e.g., Sadler et al., 2022; Zhu & Piotrowski, 2020; Zwart et al., 2021). ML models can also be run with substantially less computational expense than process models, and thus can be used to explore different scenarios for decision-making. Early stakeholder engagement can help determine desired objectives, accuracies, and lead times (Castilla-Rho, 2017; White, 2017), which could be enabled through collaborations between hydrological/water quality experts and data scientists with stakeholders such as watershed managers, regulators, and the public (Sun & Scanlon, 2019). Additionally, incorporating explainable AI approaches into decision frameworks will be important to increase transparency and build stakeholder trust in model projections. Finally, there are ethical considerations regarding the use of ML for decision-making, which include examining algorithmic assumptions, data bias and quality, and human-derived decision rules (Lo Piano, 2020).

In summary, ML paired with other emerging technologies such as cloud and exascale computing, 5G networks, and big data tools has the potential to significantly advance data-driven decisions and knowledge discovery for watershed and water quality management (Hubbard et al., 2020; Sun & Scanlon, 2019). The field of ML, and more broadly AI, has had a long history of developing other relevant technologies that are not discussed here including intelligent agents, decision theory, and reinforcement learning, which warrant further consideration in data-driven decision support systems for watershed management (Russell & Norvig, 2020). The potential for improving water quality models using ML can be realized by making deliberate choices for model selection, design, and evaluation based on the above considerations and incorporating the latest computational and mathematical advances.

5 | CONCLUSIONS

Improving water quality models for timely, decision-relevant predictions would be beneficial in the face of climate and land use change, growing populations, and a greater likelihood of extreme events. Water quality modelling poses many challenges due to the diversity of parameters and processes involved, the need for scaling, and data availability. The rapid growth in use of the state-of-the-art ML and DL models in hydrology, and recent computational and mathematical advances have demonstrated the potential for using ML for water quality modelling. Models can be improved using new approaches for model selection, hyperparameter optimization, knowledge-guided ML, transfer learning, new representations for complex data, explicit treatment for extremes, UQ, and explainable AI methods. At the same time, development of tools to integrate and process observational datasets, and methods for data assimilation would enhance the use of ML for water quality modelling. For decision-relevant predictions, consideration of stakeholder needs in model selection and evaluation of trade-offs between desired accuracy, complexity, timeliness, and available data and computational resources are important. With thoughtful implementation, ML models have the potential to accelerate decision-relevant predictions and process understanding of river water quality.

ACKNOWLEDGEMENTS

This paper is supported by the iNAIADS DOE Early Career Project and Watershed Function Science Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award no. DE-AC02-05CH11231. Juliane Müller and Talita Perciano are supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory and the Office of Advanced Scientific Computing Research under U.S. Department of Energy Contract No. DE-AC02-05CH11231. Contributions of Samantha Oliver, Jacob Zwart, and Alison Appling were supported by the Water Mission Area of the U.S. Geological Survey. Vipin Kumar and Jared Willard were supported by NSF grant#1934721. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analysed for this study.

ORCID

Charuleka Varadharajan https://orcid.org/0000-0002-4142-3224
Alison P. Appling https://orcid.org/0000-0003-3638-8572
Bhavna Arora https://orcid.org/0000-0001-7841-886X
Danielle S. Christianson https://orcid.org/0000-0002-8663-7701
Valerie C. Hendrix https://orcid.org/0000-0001-9061-8952
Vipin Kumar https://orcid.org/0000-0002-9040-2665
Aranildo R. Lima https://orcid.org/0000-0003-2514-891X
Juliane Müller https://orcid.org/0000-0001-8627-1992
Samantha Oliver https://orcid.org/0000-0001-5668-1165
Mohammed Ombadi https://orcid.org/0000-0001-7793-9137
Talita Perciano https://orcid.org/0000-0002-2388-1803
Jeffrey M Sadler https://orcid.org/0000-0001-8776-4844
Helen Weierbach https://orcid.org/0000-0001-6348-9120
Jared D. Willard https://orcid.org/0000-0003-4434-051X
Zexuan Xu https://orcid.org/0000-0001-9534-7370
Jacob Zwart https://orcid.org/0000-0002-3870-405X

REFERENCES

Abbott, B. W., Bishop, K., Zarnetske, J. P., Hannah, D. M., Frei, R. J., Minaudo, C., Chapin, F. S., Krause, S., Conner, L., Ellison, D., Godsey, S. E., Plont, S., Marçais, J., Kolbe, T., Huebner, A., Hampton, T., Gu, S., Buhman, M., Sayedi, S. S., … Pinay, G. (2019). A water cycle for the Anthropocene. *Hydrological Processes*, *33*(23), 3046–3052. https://doi.org/10.1002/hyp.13544

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, *76*, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Althoff, D., Bazame, H. C., & Nascimento, J. G. (2021). Untangling hybrid hydrological models with explainable artificial intelligence. *H2Open Journal*, *4*(1), 13–28. https://doi.org/10.2166/h2oj.2021.066

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., & Over, T. (2015). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, *51*(12), 10078–10091. https://doi.org/10.1002/2015WR017498

Álvarez-Cabria, M., Barquín, J., & Peñas, F. J. (2016). Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors. *Science of the Total Environment*, *545–546*, 152–162. https://doi.org/10.1016/j.scitotenv.2015.12.109

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment Part I: Model development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Arora, B., Burrus, M., Newcomer, M., Steefel, C. I., Carroll, R. W. H., Dwivedi, D., Dong, W., Williams, K. H., & Hubbard, S. S. (2020). Differential C-Q analysis: A new approach to inferring lateral transport and hydrologic transients within multiple reaches of a mountainous headwater catchment. *Frontiers in Water*, *2*, 24. https://doi.org/10.3389/frwa.2020.00024

Arora, B., Dwivedi, D., Hubbard, S. S., Steefel, C. I., & Williams, K. H. (2016). Identifying geochemical hot moments and their controls on a contaminated river floodplain system using wavelet and entropy approaches. *Environmental Modelling & Software*, *85*, 27–41. https://doi.org/10.1016/j.envsoft.2016.08.005

Arora, B., Mohanty, B. P., & McGuire, J. T. (2015). An integrated Markov chain Monte Carlo algorithm for upscaling hydrological and geochemical parameters from column to field scale. *The Science of the Total Environment*, *512–513*, 428–443. https://doi.org/10.1016/j.scitotenv.2015.01.048

Arora, B., Spycher, N. F., Steefel, C. I., Molins, S., Bill, M., Conrad, M. E., Dong, W., Faybishenko, B., Tokunaga, T. K., Wan, J., Williams, K. H., & Yabusaki, S. B. (2016). Influence of hydrological, biogeochemical and temperature transients on subsurface carbon fluxes in a flood plain environment. *Biogeochemistry*, *127*(2–3), 367–396. https://doi.org/10.1007/s10533-016-0186-8

Arora, B., Wainwright, H. M., Dwivedi, D., Vaughn, L. J. S., Curtis, J. B., Torn, M. S., Dafflon, B., & Hubbard, S. S. (2019). Evaluating temporal controls on greenhouse gas (GHG) fluxes in an Arctic tundra environment: An entropy-based approach. *Science of the Total Environment*, *649*, 284–299. https://doi.org/10.1016/j.scitotenv.2018.08.251

Ba, Y., Zhao, G., & Kadambi, A. (2019). Blending diverse physical priors with neural networks. *ArXiv:1910.00201 [Cs, stat]*. http://arxiv.org/abs/1910.00201

Baek, S.-S., Pyo, J., & Chun, J. A. (2020). Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water*, *12*(12), 3399. https://doi.org/10.3390/w12123399

Belitz, K., & Stackelberg, P. E. (2021). Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. *Environmental Modelling & Software*, *139*, 105006. https://doi.org/10.1016/j.envsoft.2021.105006

Bennett, A., & Nijssen, B. (2021). Deep learned process parameterizations provide better representations of turbulent heat fluxes in hydrologic models. *Water Resources Research*, *57*(5), e2020WR029328. https://doi.org/10.1029/2020WR029328

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science*, *363*(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems (Vol. 24)*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, *13*, 281–305.

Bergstra, J., Yamins, D., & Cox, D. (2013). Hyperopt: A Python Library for optimizing the hyperparameters of machine learning algorithms. Proceedings of the 12th Python in Science Conference, 13–19. https://doi.org/10.25080/Majora-8b375195-003

Bernhardt, E. S., Blaszczak, J. R., Ficken, C. D., Fork, M. L., Kaiser, K. E., & Seybold, E. C. (2017). Control points in ecosystems: Moving beyond the hot spot hot moment concept. *Ecosystems*, *20*(4), 665–682. https://doi.org/10.1007/s10021-016-0103-y

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., & Donigian, A. S. (1996). *Hydrologic simulation program—Fortran user's manual for release* (Vol. 11, p. 880) https://hydrologicmodels.tamu.edu/wp-content/uploads/sites/103/2018/09/HSPF_User-Manual.pdf

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M., & Srinivasan, R. (2017). Introduction to SWAT+, a completely restructured version of the soil and water assessment tool. *JAWRA Journal of the American Water Resources Association*, *53*(1), 115–130. https://doi.org/10.1111/1752-1688.12482

Bieroza, M. Z., Heathwaite, A. L., Bechmann, M., Kyllmar, K., & Jordan, P. (2018). The concentration-discharge slope as a tool for water quality management. *Science of the Total Environment*, *630*, 738–749. https://doi.org/10.1016/j.scitotenv.2018.02.256

Bishop, C. (2006). *Pattern recognition and machine learning* (1st ed.). Springer https://link.springer.com/book/9780387310732

Blodgett, D., Lucido, J., & Kreft, J. (2015). Progress on water data integration and distribution: A summary of select US geological survey data systems. *Journal of Hydroinformatics*, *18*(2), 226–237. https://doi.org/10.2166/hydro.2015.067

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the Lorenz 96 model. *Journal of Computational Science*, *44*, 101171. https://doi.org/10.1016/j.jocs.2020.101171

Brantley, S. L., McDowell, W. H., Dietrich, W. E., White, T. S., Kumar, P., Anderson, S. P., Chorover, J., Lohse, K. A., Bales, R. C., Richter, D. D., Grant, G., & Gaillardet, J. (2017). Designing a network of critical zone observatories to explore the living skin of the terrestrial earth. *Earth Surface Dynamics*, *5*(4), 841–860. https://doi.org/10.5194/esurf-5-841-2017

Breiman, L., & Friedman, J. H. (1985). Estimating pptimal transformations for multiple regression and correlation. *Journal of the American*

Statistical Association, *80*(391), 580–598. https://doi.org/10.1080/01621459.1985.10478157

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment*, *721*, 137612. https://doi.org/10.1016/j.scitotenv.2020.137612

da Silva Burigato Costa, C. M., da Silva Marques, L., Almeida, A. K., Leite, I. R., & de Almeida, I. K. (2019). Applicability of water quality models around the world—A review. *Environmental Science and Pollution Research*, *26*(36), 36141–36162. https://doi.org/10.1007/s11356-019-06637-2

Burns, D. A., McDonnell, T. C., Rice, K. C., Lawrence, G. B., & Sullivan, T. J. (2020). Chronic and episodic acidification of streams along the Appalachian Trail corridor, eastern United States. *Hydrological Processes*, *34*(7), 1498–1513. https://doi.org/10.1002/hyp.13668

Caissie, D. (2006). The thermal regime of rivers: A review. *Freshwater Biology*, *51*(8), 1389–1406. https://doi.org/10.1111/j.1365-2427.2006.01597.x

Campbell, J. L., Rustad, L. E., Porter, J. H., Taylor, J. R., Dereszynski, E. W., Shanley, J. B., Gries, C., Henshaw, D. L., Martin, M. E., Sheldon, W. M., & Boose, E. R. (2013). Quantity is nothing without quality: Automated QA/QC for streaming environmental sensor data. *Bioscience*, *63*(7), 574–585. https://doi.org/10.1525/bio.2013.63.7.10

Cannon, A. J. (2010). A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrological Processes*, *24*(6), 673–685. https://doi.org/10.1002/hyp.7506

Cannon, A. J. (2012). Neural networks for probabilistic environmental prediction: Conditional density estimation network creation and evaluation (CaDENCE) in R. *Computers & Geosciences*, *41*, 126–135. https://doi.org/10.1016/j.cageo.2011.08.023

Cantor, A., Kiparsky, M., Hubbard, S. S., Kennedy, R., Pecharroman, L. C., Guivetchi, K., Darling, G., McCready, C., & Bales, R. (2021). Making a water data system responsive to information needs of decision makers. *Frontiers in Climate*, *3*, 143. https://doi.org/10.3389/fclim.2021.761444

Castellini, A., Bloisi, D., Blum, J., Masillo, F., & Farinelli, A. (2020). Multivariate sensor signals collected by aquatic drones involved in water monitoring: A complete dataset. *Data in Brief*, *30*, 105436. https://doi.org/10.1016/j.dib.2020.105436

Castellini, A., Masillo, F., Bicego, M., Bloisi, D., Blum, J., Farinelli, A., & Peigner, S. (2019). Subspace clustering for situation assessment in aquatic drones. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. Association for Computing Machinery (pp. 930–937). https://doi.org/10.1145/3297280.3297372

Castilla-Rho, J. C. (2017). Groundwater modeling with stakeholders: Finding the complexity that matters. *Groundwater*, *55*(5), 620–625. https://doi.org/10.1111/gwat.12569

Cawley, G. C., Janacek, G. J., Haylock, M. R., & Dorling, S. R. (2007). Predictive uncertainty in environmental modelling. *Neural Networks*, *20*(4), 537–549. https://doi.org/10.1016/j.neunet.2007.04.024

Chanat, J. G., & Yang, G. (2018). Exploring drivers of regional water-quality change using differential spatially referenced regression—A pilot study in the Chesapeake Bay watershed. *Water Resources Research*, *54*(10), 8120–8145. https://doi.org/10.1029/2017WR022403

Chen, P.-H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. *Nature Materials*, *18*(5), 410–414. https://doi.org/10.1038/s41563-019-0345-0

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2019). Neural ordinary differential equations. *ArXiv:1806.07366 [Cs, stat]*. http://arxiv.org/abs/1806.07366

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, *10*(17), 5776. https://doi.org/10.3390/app10175776

Chen, Y., Xu, C.-Y., Chen, X., Xu, Y., Yin, Y., Gao, L., & Liu, M. (2019). Uncertainty in simulation of land-use change impacts on catchment runoff with multi-timescales based on the comparison of the HSPF and SWAT models. *Journal of Hydrology*, *573*, 486–500. https://doi.org/10.1016/j.jhydrol.2019.03.091

Coulibaly, P., Bobée, B., & Anctil, F. (2001). Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. *Hydrological Processes*, *15*(8), 1533–1536. https://doi.org/10.1002/hyp.445

Culhane, P. J., Friesema, H. P., & Beecher, J. A. (1987). *Project impacts: The accuracy of individual forecasts*. CRC Press.

Danandeh Mehr, A., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M. A., & Yaseen, Z. M. (2018). Genetic programming in water resources engineering: A state-of-the-art review. *Journal of Hydrology*, *566*, 643–667. https://doi.org/10.1016/j.jhydrol.2018.09.043

Daw, A., Thomas, R. Q., Carey, C. C., Read, J. S., Appling, A. P., & Karpatne, A. (2020). Physics-guided architecture (PGA) of neural networks for quantifying uncertainty in lake temperature modeling. In *Proceedings of the 2020 SIAM International Conference on Data Mining (SDM)* (pp. 532–540). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611976236.60

de Oliviera, C. F. (2019). *Metalearning for multiple-domain transfer learning*. https://www.proquest.com/openview/987b2fd7140316c9ae9e2d47bab1436e/1?pq-origsite=gscholar&cbl=51922&diss=y

DeFlorio, M. J., Ralph, F. M., Waliser, D. E., Jones, J., & Anderson, M. L. (2021). *Better subseasonal-to-seasonal forecasts for water management*. Eos http://eos.org/features/better-subseasonal-to-seasonal-forecasts-for-water-management

Duan, S., Ullrich, P., & Shu, L. (2020). Using convolutional neural networks for streamflow projection in California. *Frontiers in Water*, *2*, 28. https://doi.org/10.3389/frwa.2020.00028

Dumont, V., Garner, C., Trivedi, A., Jones, C., Ganapati, V., Mueller, J., Perciano, T., Kiran, M., & Day, M. (2021). HYPPO: A surrogate-based multi-level parallelism tool for hyperparameter optimization. *ArXiv: 2110.01698 [Cs]*. http://arxiv.org/abs/2110.01698

Dwivedi, D., Steefel, C. I., Arora, B., Newcomer, M., Moulton, J. D., Dafflon, B., Faybishenko, B., Fox, P., Nico, P., Spycher, N., Carroll, R., & Williams, K. H. (2018). Geochemical exports to river from the intrameander hyporheic zone under transient hydrologic conditions: East River mountainous watershed, Colorado. *Water Resources Research*, *54*(10), 8456–8477. https://doi.org/10.1029/2018WR023377

Evans, C., & Davies, T. D. (1998). Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry. *Water Resources Research*, *34*(1), 129–137. https://doi.org/10.1029/97WR01881

Fallah-Mehdipour, E., & Haddad, O. B. (2015). Application of genetic programming in hydrology. In A. H. Gandomi, A. H. Alavi, & C. Ryan (Eds.), *Handbook of genetic programming applications* (pp. 59–70). Springer International Publishing. https://doi.org/10.1007/978-3-319-20883-1_3

Feigl, M., Lebiedzinski, K., Herrnegger, M., & Schulz, K. (2021). Machine-learning methods for stream water temperature prediction. *Hydrology and Earth System Sciences*, *25*(5), 2951–2977. https://doi.org/10.5194/hess-25-2951-2021

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. *Advances in Neural Information Processing Systems*, *28*, 28. https://proceedings.neurips.cc/paper/2015/hash/11d0e6287202fced83f79975ec59a3a6-Abstract.html

Fleming, S. W., Bourdin, D. R., Campbell, D., Stull, R. B., & Gardner, T. (2015). Development and operational testing of a super-ensemble artificial intelligence flood-forecast model for a Pacific Northwest River. *JAWRA Journal of the American Water Resources Association*, *51*(2), 502–512. https://doi.org/10.1111/jawr.12259

Fleming, S. W., Garen, D. C., Goodbody, A. G., McCarthy, C. S., & Landers, L. C. (2021). 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*, 126782. https://doi.org/10.1016/j.jhydrol.2021.126782

Forio, M. A. E., Landuyt, D., Bennetsen, E., Lock, K., Nguyen, T. H. T., Ambarita, M. N. D., Musonge, P. L. S., Boets, P., Everaert, G., Dominguez-Granda, L., & Goethals, P. L. M. (2015). Bayesian belief network models to analyse and predict ecological water quality in rivers. *Ecological Modelling*, *312*, 222–238. https://doi.org/10.1016/j.ecolmodel.2015.05.025

Frame, J., Nearing, G., Kratzert, F., & Rahman, M. (2021). Post-processing the National Water Model with long short-term memory networks for streamflow predictions and model diagnostics. https://eartharxiv.org/repository/view/124/

Frame, J., Kratzert, F., Klotz, D., Gauch, M., Shelev, G., Gilon, O., Qualls, L. M., Gupta, H. V., & Nearing, G. S. (2021). Deep learning rainfall-runoff predictions of extreme events. *Hydrology and Earth System Sciences Discussions*, *1–20*. https://doi.org/10.5194/hess-2021-423

Fu, B., Horsburgh, J. S., Jakeman, A. J., Gualtieri, C., Arnold, T., Marshall, L., Green, T. R., Quinn, N. W. T., Volk, M., Hunt, R. J., Vezzaro, L., Croke, B. F. W., Jakeman, J. D., Snow, V., & Rashleigh, B. (2020). Modeling water quality in watersheds: From here to the next generation. *Water Resources Research*, *56*(11), e2020WR027721. https://doi.org/10.1029/2020WR027721

Fu, B., Merritt, W. S., Croke, B. F. W., Weber, T. R., & Jakeman, A. J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software*, *114*, 75–97. https://doi.org/10.1016/j.envsoft.2018.12.008

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*. Proceedings of Machine Learning Research. (Vol. 48, pp. 1050–1059).

Gast, J., & Roth, S. (2018). *Lightweight probabilistic deep networks* (pp. 3369–3378). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://openaccess.thecvf.com/content_cvpr_2018/html/Gast_Lightweight_Probabilistic_Deep_CVPR_2018_paper.html

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., & Hochreiter, S. (2021). Rainfall–runoff prediction at multiple timescales with a single long short-term memory network. *Hydrology and Earth System Sciences*, *25*(4), 2045–2062. https://doi.org/10.5194/hess-25-2045-2021

Gharib, A., & Davies, E. G. R. (2021). A workflow to address pitfalls and challenges in applying machine learning models to hydrology. *Advances in Water Resources*, *152*, 103920. https://doi.org/10.1016/j.advwatres.2021.103920

Godsey, S. E., Kirchner, J. W., & Clow, D. W. (2009). Concentration–discharge relationships reflect chemostatic characteristics of US catchments. *Hydrological Processes*, *23*(13), 1844–1864. https://doi.org/10.1002/hyp.7315

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Green, M. B., Pardo, L. H., Bailey, S. W., Campbell, J. L., McDowell, W. H., Bernhardt, E. S., & Rosi, E. J. (2021). Predicting high-frequency variation in stream solute concentrations with water quality sensors and machine learning. *Hydrological Processes*, *35*(1), e14000. https://doi.org/10.1002/hyp.14000

Guo, D., Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Leahy, P., Wilson, P., & Western, A. W. (2019). Key factors affecting temporal variability in stream water quality. *Water Resources Research*, *55*(1), 112–129. https://doi.org/10.1029/2018WR023370

Hamshaw, S. D., Denu, D., Holthuijzen, M., Wshah, S., & Rizzo, D. (2019, 2019). *Automating the classification of hysteresis in event concentration-discharge relationships*. SEDHYD https://www.sedhyd.org/2019/openconf/modules/request.php?module=oc_program&action=view.php&id=70&file=1/70.pdf

Hamshaw, S. D., Dewoolkar, M. M., Schroth, A. W., Wemple, B. C., & Rizzo, D. M. (2018). A new machine-learning approach for classifying hysteresis in suspended-sediment discharge relationships using high-frequency monitoring data. *Water Resources Research*, *54*(6), 4040–4058. https://doi.org/10.1029/2017WR022238

Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P., & Ghil, M. (2016). Causal counterfactual theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society*, *97*(1), 99–110. https://doi.org/10.1175/BAMS-D-14-00034.1

Hanson, P. C., Stillman, A. B., Jia, X., Karpatne, A., Dugan, H. A., Carey, C. C., Stachelek, J., Ward, N. K., Zhang, Y., Read, J. S., & Kumar, V. (2020). Predicting lake surface water phosphorus dynamics using process-guided machine learning. *Ecological Modelling*, *430*, 109136. https://doi.org/10.1016/j.ecolmodel.2020.109136

Harmel, R. D., Smith, P. K., Migliaccio, K. W., Chaubey, I., Douglas-Mankin, K. R., Benham, B., Shukla, S., Muñoz-Carpena, R., & Robson, B. J. (2014). Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. *Environmental Modelling & Software*, *57*, 40–51. https://doi.org/10.1016/j.envsoft.2014.02.013

Harrison, J. W., Lucius, M. A., Farrell, J. L., Eichler, L. W., & Relyea, R. A. (2021). Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using random forests regression. *Science of the Total Environment*, *763*, 143005. https://doi.org/10.1016/j.scitotenv.2020.143005

Hartmann, J., Moosdorf, N., Lauerwald, R., Hinderer, M., & West, A. J. (2014). Global chemical weathering and associated P-release—The role of lithology, temperature and soil properties. *Chemical Geology*, *363*, 145–163. https://doi.org/10.1016/j.chemgeo.2013.10.025

Hassan, N., & Woo, C. S. (2021). Machine learning application in water quality using satellite data. *IOP Conference Series: Earth and Environmental Science*, *842*(1), 012018. https://doi.org/10.1088/1755-1315/842/1/012018

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

He, Q., Barajas-Solano, D., Tartakovsky, G., & Tartakovsky, A. M. (2020). Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources*, *141*, 103610. https://doi.org/10.1016/j.advwatres.2020.103610

Her, Y., Frankenberger, J., Chaubey, I., & Srinivasan, R. (2015). Threshold effects in HRU definition of the soil and water assessment tool. *Transactions of the ASABE*, *58*(2), 367–378. https://doi.org/10.13031/trans.58.10805

Herath, H. M. V. V., Chadalawada, J., & Babovic, V. (2021). Genetic programming for hydrological applications: To model or to forecast that is the question. *Journal of Hydroinformatics*, *23*(4), 740–763. https://doi.org/10.2166/hydro.2021.179

Hestir, E. L., Brando, V. E., Bresciani, M., Giardino, C., Matta, E., Villa, P., & Dekker, A. G. (2015). Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission. *Remote Sensing of Environment*, *167*, 181–195. https://doi.org/10.1016/j.rse.2015.05.023

Hill, D. J., & Minsker, B. S. (2010). Anomaly detection in streaming environmental sensor data: A data-driven modeling approach. *Environmental Modelling & Software*, *25*(9), 1014–1022. https://doi.org/10.1016/j.envsoft.2009.08.010

Hill, P. R., Kumar, A., Temimi, M., & Bull, D. R. (2020). HABNet: Machine learning, remote sensing-based detection of harmful algal blooms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 3229–3239. https://doi.org/10.1109/JSTARS.2020.3001445

Horsburgh, J. S., Morsy, M. M., Castronova, A. M., Goodall, J. L., Gan, T., Yi, H., Stealey, M. J., & Tarboton, D. G. (2016). HydroShare: Sharing diverse environmental data types and models as social objects with application to the hydrology domain. *JAWRA Journal of the American Water Resources Association*, *52*(4), 873–889. https://doi.org/10.1111/1752-1688.12363

Hoyer, P., Janzing, D., Mooij, J. M., Peters, J., & Schölkopf, B. (2009). Nonlinear causal discovery with additive noise models. *Advances in Neural Information Processing Systems*, 21 https://papers.nips.cc/paper/2008/hash/f7664060cc52bc6f3d620bcedc94a4b6-Abstract.html

Hsieh, W. W. (2009). *Machine learning methods in the environmental sciences: Neural networks and kernels*. Cambridge University Press. https://doi.org/10.1017/CBO9780511627217

Hu, X., Chu, L., Pei, J., Liu, W., & Bian, J. (2021). Model complexity of deep learning: A survey. *ArXiv:2103.05127 [Cs]*. http://arxiv.org/abs/2103.05127

Hubbard, S. S., Varadharajan, C., Wu, Y., Wainwright, H., & Dwivedi, D. (2020). Emerging technologies and radical collaboration to advance predictive understanding of watershed hydrobiogeochemistry. *Hydrological Processes*, *34*, 3175–3182. https://doi.org/10.1002/hyp.13807

Hunter, J. M., Maier, H. R., Gibbs, M. S., Foale, E. R., Grosvenor, N. A., Harders, N. P., & Kikuchi-Miller, T. C. (2018). Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems. *Hydrology and Earth System Sciences*, *22*(5), 2987–3006. https://doi.org/10.5194/hess-22-2987-2018

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). *Automated machine learning: Methods, systems, challenges*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5

Jan, A., Coon, E. T., & Painter, S. L. (2021). Toward more mechanistic representations of biogeochemical processes in river networks: Implementation and demonstration of a multiscale model. *Environmental Modelling & Software*, *145*, 105166. https://doi.org/10.1016/j.envsoft.2021.105166

Javed, A., Hamshaw, S. D., Lee, B. S., & Rizzo, D. M. (2021). Multivariate event time series analysis using hydrological and suspended sediment data. *Journal of Hydrology*, *593*, 125802. https://doi.org/10.1016/j.jhydrol.2020.125802

Jia, X., Lin, B., Zwart, J., Sadler, J., Appling, A., Oliver, S., & Read, J. (2021). Graph-based reinforcement learning for active learning in real time: An application in modeling river networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*. Society for Industrial and Applied Mathematics (pp. 621–629). https://doi.org/10.1137/1.9781611976700.70

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. *ArXiv:1810.13075 [Physics]*. http://arxiv.org/abs/1810.13075

Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., Willard, J., Xu, S., Steinbach, M., Read, J., & Kumar, V. (2021). Physics-guided recurrent graph model for predicting flow and temperature in river networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)* (pp. 612–620). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611976700.69

Jiang, S., Ren, L., Yang, X., Ma, M., & Liu, Y. (2014). Multi-model ensemble hydrologic prediction and uncertainties analysis. *Proceedings of the International Association of Hydrological Sciences*, *364*, 249–254. https://doi.org/10.5194/piahs-364-249-2014

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*, 255–260. https://doi.org/10.1126/science.aaa8415

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2018). Physics-guided neural networks (PGNN): An application in lake temperature modeling. *ArXiv:1710.11431 [Physics, Stat]*. http://arxiv.org/abs/1710.11431

Kashinath, K., Mustafa, M., Albert, A., Wu, J.-L., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., Singh, A., Manepalli, A., Chirila, D., Yu, R., Walters, R., White, B., Xiao, H., Tchelepi, H. A., Marcus, P., Anandkumar, A., … Prabhat. (2021). Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2194), 20200093. https://doi.org/10.1098/rsta.2020.0093

Kaushal, S. S., Likens, G. E., Pace, M. L., Reimer, J. E., Maas, C. M., Galella, J. G., Utz, R. M., Duan, S., Kryger, J. R., Yaculak, A. M., Boger, W. L., Bailey, N. W., Haq, S., Wood, K. L., Wessel, B. M., Park, C. E., Collison, D. C., Aisin, B. Y., Aaqob, I., … Woglo, S. A. (2021). Freshwater salinization syndrome: From emerging global problem to managing risks. *Biogeochemistry*, *154*(2), 255–292. https://doi.org/10.1007/s10533-021-00784-w

Kaushal, S. S., Likens, G. E., Pace, M. L., Utz, R. M., Haq, S., Gorman, J., & Grese, M. (2018). Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(4), E574–E583. https://doi.org/10.1073/pnas.1711234115

Kaushal, S. S., Likens, G. E., Utz, R. M., Pace, M. L., Grese, M., & Yepsen, M. (2013). Increased river alkalinization in the eastern U.S. *Environmental Science & Technology*, *47*(18), 10302–10311. https://doi.org/10.1021/es401046s

Keiser, D. A., Kling, C. L., & Shapiro, J. S. (2019). The low but uncertain measured benefits of US water quality policy. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5262–5269. https://doi.org/10.1073/pnas.1802870115

Khandelwal, A., Xu, S., Li, X., Jia, X., Stienbach, M., Duffy, C., Nieber, J., & Kumar, V. (2020). Physics guided machine learning methods for hydrology. *ArXiv:2012.02854 [Physics]*. http://arxiv.org/abs/2012.02854

Khullar, S., & Singh, N. (2020). Machine learning techniques in river water quality modelling: A research travelogue. *Water Supply*, *21*(1), 1–13. https://doi.org/10.2166/ws.2020.277

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.

Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Kratzert, F., Klotz, D., Brandstetter, J., Hoedt, P.-J., Nearing, G., & Hochreiter, S. (2019). Using LSTMs for climate change assessment studies on droughts and floods. *ArXiv:1911.03941 [Physics, Stat]*. http://arxiv.org/abs/1911.03941

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall—Runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354. https://doi.org/10.1029/2019WR026065

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kruse, P. (2018). Review on water quality sensors. *Journal of Physics D: Applied Physics*, *51*(20), 203002. https://doi.org/10.1088/1361-6463/aabb93

Kunin, D., Sagastuy-Brena, J., Ganguli, S., Yamins, D. L. K., & Tanaka, H. (2021). Neural mechanics: Symmetry and broken conservation laws in deep learning dynamics. *ArXiv:2012.04728 [Cond-Mat, q-Bio, Stat]*. http://arxiv.org/abs/2012.04728

Larsen, S., Hamilton, S., Lucido, J., Garner, B., & Young, D. (2016). Supporting diverse data providers in the open water data initiative: Communicating water data quality and fitness of use. *JAWRA Journal of the American Water Resources Association*, *52*(4), 859–872. https://doi.org/10.1111/1752-1688.12406

Laude, E., Lange, J., Schüpfer, J., Domokos, C., Leal-Taixé, L., Schmidt, F. R., Andres, B., & Cremers, D. (2018). Discrete-continuous ADMM for transductive inference in higher-order MRFs. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/CVPR.2018.00174

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, S., Kim, J., Hwang, J., Lee, E., Lee, K.-J., Oh, J., Park, J., & Heo, T.-Y. (2020). Clustering of time series water quality data using dynamic time warping: A case study from the Bukhan river water quality monitoring network. *Water*, *12*(9), 2411. https://doi.org/10.3390/w12092411

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., Kumar Sahu, R., Greve, P., Slater, L., & Dadson, S. (2021). Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences Discussions*, 1–37. https://doi.org/10.5194/hess-2021-566

Leigh, C., Alsibai, O., Hyndman, R. J., Kandanaarachchi, S., King, O. C., McGree, J. M., Neelamraju, C., Strauss, J., Talagala, P. D., Turner, R. D. R., Mengersen, K., & Peterson, E. E. (2019). A framework for automated anomaly detection in high frequency water-quality data from in situ sensors. *Science of the Total Environment*, *664*, 885–898. https://doi.org/10.1016/j.scitotenv.2019.02.085

Li, H.-Y., Leung, L. R., Tesfa, T., Voisin, N., Hejazi, M., Liu, L., Liu, Y., Rice, J., Wu, H., & Yang, X. (2015). Modeling stream temperature in the Anthropocene: An earth system modeling approach. *Journal of Advances in Modeling Earth Systems*, *7*(4), 1661–1679.

Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., Moore, J., Perdrial, J., Sullivan, P., Thompson, A., Jin, L., Bolton, E. W., Brantley, S. L., Dietrich, W. E., Mayer, K. U., Steefel, C. I., Valocchi, A., Zachara, J., Kocar, B., … Beisman, J. (2017). Expanding the role of reactive transport models in critical zone processes. *Earth-Science Reviews*, *165*, 280–301. https://doi.org/10.1016/j.earscirev.2016.09.001

Li, L., Sullivan, P. L., Benettin, P., Cirpka, O. A., Bishop, K., Brantley, S. L., Knapp, J. L. A., van Meerveld, I., Rinaldo, A., Seibert, J., Wen, H., & Kirchner, J. W. (2021). Toward catchment hydro-biogeochemical theories. *WIREs Water*, *8*(1), e1495. https://doi.org/10.1002/wat2.1495

Li, T., Zhang, J., Bao, K., Liang, Y., Li, Y., & Zheng, Y. (2020). AutoST: Efficient neural architecture search for spatio-temporal prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 794–802). Association for Computing Machinery. https://doi.org/10.1145/3394486.3403122

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674. https://doi.org/10.3390/s18082674

Ling, J., Kurzawski, A., & Templeton, J. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, *807*, 155–166. https://doi.org/10.1017/jfm.2016.615

Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Waters, D., Leahy, P., Wilson, P., & Western, A. W. (2018a). Key factors influencing differences in stream water quality across space. *Wiley Interdisciplinary Reviews: Water*, *5*(1), e1260. https://doi.org/10.1002/wat2.1260

Lintern, A., Webb, J. A., Ryu, D., Liu, S., Waters, D., Leahy, P., Bende-Michl, U., & Western, A. W. (2018b). What are the key catchment characteristics affecting spatial differences in riverine water quality? *Water Resources Research*, *54*(10), 7252–7272. https://doi.org/10.1029/2017WR022172

Liu, J., Zhang, D., Tang, Q., Xu, H., Huang, S., Shang, D., & Liu, R. (2021). Water quality assessment and source identification of the Shuangji River (China) using multivariate statistical methods. *PLoS One*, *16*(1), e0245525. https://doi.org/10.1371/journal.pone.0245525

Liu, P., Wang, J., Sangaiah, A. K., Xie, Y., & Yin, X. (2019). Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability*, *11*(7), 2058. https://doi.org/10.3390/su11072058

Liu, W., Birgand, F., Tian, S., & Chen, C. (2021). Event-scale hysteresis metrics to reveal processes and mechanisms controlling constituent export from watersheds: A review☆. *Water Research*, *200*, 117254. https://doi.org/10.1016/j.watres.2021.117254

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 1–7. https://doi.org/10.1057/s41599-020-0501-9

Loquercio, A., Segu, M., & Scaramuzza, D. (2020). A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, *5*(2), 3153–3160. https://doi.org/10.1109/LRA.2020.2974682

Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, *249*, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *ArXiv:1705.07874 [Cs, stat]*. http://arxiv.org/abs/1705.07874

Lyubimova, T., Lepikhin, A., Parshakova, Y., & Tiunov, A. (2016). The risk of river pollution due to washout from contaminated floodplain water bodies during periods of high magnitude floods. *Journal of Hydrology*, *534*, 579–589. https://doi.org/10.1016/j.jhydrol.2016.01.030

Ma, B., Guo, W., & Zhang, J. (2020). A survey of online data-driven proactive 5G network optimisation using machine learning. *IEEE Access*, *8*, 35606–35637. https://doi.org/10.1109/ACCESS.2020.2975004

McDowell, R. W., Noble, A., Pletnyakov, P., Haggard, B. E., & Mosley, L. M. (2020). Global mapping of freshwater nutrient enrichment and periphyton growth potential. *Scientific Reports*, *10*(1), 3568. https://doi.org/10.1038/s41598-020-60279-w

Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A., & Pham, B. T. (2020). River water salinity prediction using hybrid machine learning models. *Water*, *12*(10), 2951. https://doi.org/10.3390/w12102951

Mesbah, S. M., Kerachian, R., & Nikoo, M. R. (2009). Developing real time operating rules for trading discharge permits in rivers: Application of Bayesian networks. *Environmental Modelling and Software*, *24*(2), 238–246. https://doi.org/10.1016/j.envsoft.2008.06.007

Mirzaei, M., Jafari, A., Gholamalifard, M., Azadi, H., Shooshtari, S. J., Moghaddam, S. M., Gebrehiwot, K., & Witlox, F. (2020). Mitigating environmental risks: Modeling the interaction of water quality parameters and land use cover. *Land Use Policy*, *95*, 103766. https://doi.org/10.1016/j.landusepol.2018.12.014

Mital, U., Dwivedi, D., Brown, J. B., Faybishenko, B., Painter, S. L., & Steefel, C. I. (2020). Sequential imputation of missing spatio-temporal precipitation data using random forests. *Frontiers in Water*, *2*, 20. https://doi.org/10.3389/frwa.2020.00020

Moatar, F., Abbott, B. W., Minaudo, C., Curie, F., & Pinay, G. (2017). Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. *Water Resources Research*, *53*(2), 1270–1287. https://doi.org/10.1002/2016WR019635

Monteith, D. T., Stoddard, J. L., Evans, C. D., de Wit, H. A., Forsius, M., Høgåsen, T., Wilander, A., Skjelkvåle, B. L., Jeffries, D. S., Vuorenmaa, J., Keller, B., Kopácek, J., & Vesely, J. (2007). Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. *Nature*, *450*(7169), 537–540. https://doi.org/10.1038/nature06316

Morton, R., & Henderson, B. L. (2008). Estimation of nonlinear trends in water quality: An improved approach using generalized additive

models. *Water Resources Research*, *44*(7), W07420. https://doi.org/10.1029/2007WR006191

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Mosley, L. M. (2015). Drought impacts on the water quality of freshwater systems; review and integration. *Earth-Science Reviews*, *140*, 203–214. https://doi.org/10.1016/j.earscirev.2014.11.010

Mount, N., & Stott, T. (2008). A discrete Bayesian network to investigate suspended sediment concentrations in an alpine proglacial zone. *Hydrological Processes*, *22*(18), 3772–3784. https://doi.org/10.1002/hyp.6981

Müller, J., Park, J., Sahu, R., Varadharajan, C., Arora, B., Faybishenko, B., & Agarwal, D. (2020). Surrogate optimization of deep neural networks for groundwater predictions. *Journal of Global Optimization*, *81*, 203–231. https://doi.org/10.1007/s10898-020-00912-0

Murdoch, P. S., Baron, J. S., & Miller, T. L. (2000). Potential effects of climate change on surface-water quality in North America. *JAWRA Journal of the American Water Resources Association*, *36*(2), 347–366. https://doi.org/10.1111/j.1752-1688.2000.tb04273.x

Murphy, J., & Sprague, L. (2019). Water-quality trends in US rivers: Exploring effects from streamflow trends and changes in watershed management. *Science of the Total Environment*, *656*, 645–658. https://doi.org/10.1016/j.scitotenv.2018.11.255

Murphy, J. C. (2020). Changing suspended sediment in United States rivers and streams: Linking sediment trends to changes in land use/cover, hydrology and climate. *Hydrology and Earth System Sciences*, *24*(2), 991–1010. https://doi.org/10.5194/hess-24-991-2020

Murray, S., Ghazali, M., & McBean, E. A. (2012). Real-time water quality monitoring: Assessment of multisensor data using Bayesian belief networks. *Journal of Water Resources Planning and Management*, *138*(1), 63–70. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000163

Musolff, A., Zhan, Q., Dupas, R., Minaudo, C., Fleckenstein, J. H., Rode, M., Dehaspe, J., & Rinke, K. (2021). Spatial and temporal variability in concentration-discharge relationships at the event scale. *Water Resources Research*, *57*(10), e2020WR029442. https://doi.org/10.1029/2020WR029442

Nearing, G. S., Klotz, D., Sampson, A. K., Kratzert, F., Gauch, M., Frame, J. M., Shalev, G., & Nevo, S. (2021). Technical note: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks. *Hydrology and Earth System Sciences Discussions*, 1–25. https://doi.org/10.5194/hess-2021-515

Nearing, G. S., Klotz, D., Sampson, A. K., Kratzert, F., Gauch, M., Frame, J. M., Shalev, G., and Nevo, S. (2021). Technical Note: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2021-515, in review, 2021

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What role does hydrological science play in the age of machine learning? *Water Resources Research*, *57*(3), e2020WR028091. https://doi.org/10.1029/2020WR028091

Newcomer, M. E., Bouskill, N. J., Wainwright, H., Maavara, T., Arora, B., Siirila-Woodburn, E. R., Dwivedi, D., Williams, K. H., Steefel, C., & Hubbard, S. S. (2021). Hysteresis patterns of watershed nitrogen retention and loss over the past 50 years in United States hydrological basins. *Global Biogeochemical Cycles*, *35*(4), e2020GB006777. https://doi.org/10.1029/2020GB006777

Nilsson, C., & Renofalt, B. M. (2008). Linking flow regime and water quality in Rivers: A challenge to adaptive catchment management. *Ecology and Society*, *13*(2), 18. https://doi.org/10.5751/ES-02588-130218

Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7), e2020WR027251. https://doi.org/10.1029/2020WR027251

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., & Zappa, M. (2015). Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, *523*, 147–159. https://doi.org/10.1016/j.jhydrol.2015.01.044

Paepae, T., Bokoro, P. N., & Kyamakya, K. (2021). From fully physical to virtual sensing for water quality assessment: A comprehensive review of the relevant state-of-the-art. *Sensors*, *21*(21), 6971. https://doi.org/10.3390/s21216971

Park, J., Lee, H., Park, C. Y., Hasan, S., Heo, T.-Y., & Lee, W. H. (2019). Algal morphological identification in watersheds for drinking water supply using neural architecture search for convolutional neural network. *Water*, *11*(7), 1338. https://doi.org/10.3390/w11071338

Park, N., Kim, M., Hoai, N. X., I R. McKay, & Kim, D.-K. (2021). Knowledge-guided dynamic systems modeling: A case study on modeling river water quality. *ArXiv:2103.00792 [Cs]*. http://arxiv.org/abs/2103.00792

Parmar, K. S., & Bhardwaj, R. (2013). Wavelet and statistical analysis of river water quality parameters. *Applied Mathematics and Computation*, *219*(20), 10172–10182. https://doi.org/10.1016/j.amc.2013.03.109

Paul, P. K., Zhang, Y., Mishra, A., Panigrahy, N., & Singh, R. (2019). Comparative study of two state-of-the-art semi-distributed hydrological models. *Water*, *11*(5), 871. https://doi.org/10.3390/w11050871

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. MIT Press.

Peterson, K. T., Sagan, V., & Sloan, J. J. (2020). Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. *GIScience & Remote Sensing*, *57*(4), 510–525. https://doi.org/10.1080/15481603.2020.1738061

Phan, T. D., Smart, J. C. R., Stewart-Koster, B., Sahin, O., Hadwen, W. L., Dinh, L. T., Tahmasbian, I., & Capon, S. J. (2019). Applications of Bayesian networks as decision support tools for water resource management under climate change and socio-economic stressors: A critical appraisal. *Water*, *11*(12), 2642. https://doi.org/10.3390/w11122642

Pohle, I., Baggaley, N., Palarea-Albaladejo, J., Stutter, M., & Glendell, M. (2021). A framework for assessing concentration-discharge catchment behavior from low-frequency water quality data. *Water Resources Research*, *57*(9), e2021WR029692. https://doi.org/10.1029/2021WR029692

Povak, N. A., Hessburg, P. F., McDonnell, T. C., Reynolds, K. M., Sullivan, T. J., Salter, R. B., & Cosby, B. J. (2014). Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA. *Water Resources Research*, *50*(4), 2798–2814. https://doi.org/10.1002/2013WR014203

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Qi, D., & Majda, A. J. (2020). Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(1), 52–59. https://doi.org/10.1073/pnas.1917285117

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., & Edelman, A. (2020). Universal differential equations for scientific machine learning. *ArXiv:2001.04385 [Cs, Math, q-Bio, Stat]*. http://arxiv.org/abs/2001.04385

Rahmani, F., Lawson, K., Ouyang, W., Appling, A., Oliver, S., & Shen, C. (2021). Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. *Environmental Research Letters*, *16*(2). https://doi.org/10.1088/1748-9326/abd501

Rahmani, F., Shen, C., Oliver, S., Lawson, K., & Appling, A. (2021). Deep learning approaches for improving prediction of daily stream temperature in data-scarce, unmonitored, and dammed basins. *Hydrological Processes*, *35*(11), e14400. https://doi.org/10.1002/hyp.14400

Rajaee, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems*, *200*, 103978. https://doi.org/10.1016/j.chemolab.2020.103978

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., Karpatne, A., Hansen, G. J. A., Hanson, P. C., Watkins, W., Steinbach, M., & Kumar, V. (2017). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019WR024922

Read, E. K., Carr, L., Cicco, L. D., Dugan, H. A., Hanson, P. C., Hart, J. A., Kreft, J., Read, J. S., & Winslow, L. A. (2017). Water quality data for national-scale aquatic research: The Water Quality Portal. *Water Resources Research*, *53*(2), 1735–1745. https://doi.org/10.1002/2016WR019993

Rehana, S. (2019). River water temperature modelling under climate change using support vector regression. In S. K. Singh & C. T. Dhanya (Eds.), *Hydrology in a changing world: Challenges in modeling* (pp. 171–183). Springer International Publishing. https://doi.org/10.1007/978-3-030-02197-9_8

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ren, P., Xiao, Y., Chang, X., Huang, P., Li, Z., Chen, X., & Wang, X. (2021). A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Computing Surveys*, *54*(4), 1–34. https://doi.org/10.1145/3447582

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. Association for Computing Machinery. (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., van Griensven, A., & van der Zee, S. E. A. T. M. (2010). New challenges in integrated water quality modelling. *Hydrological Processes*, *24*(24), 3447–3461. https://doi.org/10.1002/hyp.7766

Roelandt, C., Goddéris, Y., Bonnet, M.-P., & Sondag, F. (2010). Coupled modeling of biospheric and chemical weathering processes at the continental scale. *Global Biogeochemical Cycles*, *24*(2), GB2004. https://doi.org/10.1029/2008GB003420

Ross, M. R. V., Topp, S. N., Appling, A. P., Yang, X., Kuhn, C., Butman, D., Simard, M., & Pavelsky, T. M. (2019). AquaSat: A data set to enable remote sensing of water quality for inland waters. *Water Resources Research*, *55*(11), 10012–10025. https://doi.org/10.1029/2019WR024883

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11), eaau4996. https://doi.org/10.1126/sciadv.aau4996

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson https://www.pearson.com/content/one-dot-com/one-dot-com/us/en/higher-education/program.html

Russo, S., Lürig, M., Hao, W., Matthews, B., & Villez, K. (2020). Active learning for anomaly detection in environmental data. *Environmental Modelling & Software*, *134*, 104869. https://doi.org/10.1016/j.envsoft.2020.104869

Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J. A., & Kumar, V. (2021). *Multi-task deep learning of daily streamflow and water temperature*. https://eartharxiv.org/repository/view/2377/

Sadler, J. M., Appling, A. P., Read, J. S., Oliver, S. K., Jia, X., Zwart, J. A., & Kumar, V. (2022). Multi-Task Deep Learning of Daily Streamflow and Water Temperature. *Water Resources Research*, *58*(4), e2021WR030138. https://doi.org/10.1029/2021WR030138

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, *109*(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483

Sanders, M. J., Markstrom, S. L., Regan, R. S., & Atkinson, R. D. (2017). Documentation of a daily mean stream temperature module—An enhancement to the precipitation-runoff modeling system. In *Documentation of a daily mean stream temperature module—An enhancement to the precipitation-runoff modeling system* (USGS Numbered Series No. 6-D4; techniques and methods) (Vol. 6-D4, p. 28). U.S. Geological Survey. https://doi.org/10.3133/tm6D4

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Schwarz, G., Hoos, A. B., Alexander, R. B., & Smith, R. A. (2006). Section 3. The SPARROW surface water-quality model—Theory, application and user documentation. In *Section 3. The SPARROW Surface Water-Quality Model—Theory, application and user documentation* (USGS numbered series no. 6-B3; techniques and methods (Vol. 6-B3). U.S. Geological Survey. https://doi.org/10.3133/tm6B3

Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., von Gunten, U., & Wehrli, B. (2010). Global water pollution and human health. *Annual Review of Environment and Resources*, *35*(1), 109–136. https://doi.org/10.1146/annurev-environ-100809-125342

Sear, D. A., Armitage, P. D., & Dawson, F. H. (1999). Groundwater dominated rivers. *Hydrological Processes*, *13*(3), 255–276. https://doi.org/10.1002/(SICI)1099-1085(19990228)13:3<255::AID-HYP737>3.0.CO;2-Y

Sene, K. (2016). *Hydrometeorology: Forecasting and applications* (2nd ed.). Springer International Publishing. https://doi.org/10.1007/978-3-319-23546-2

Shaughnessy, A. R., Wen, T., Niu, X., & Brantley, S. L. (2019). Three Principles to Use in Streamlining Water Quality Research through Data Uniformity. *Environmental Science & Technology*, *53*(23), 13549–13550. https://doi.org/10.1021/acs.est.9b06406

Shaughnessy, A. R., Gu, X., Wen, T., & Brantley, S. L. (2021). Machine learning deciphers CO_2 sequestration and subsurface flowpaths from stream chemistry. *Hydrology and Earth System Sciences*, *25*(6), 3397–3409. https://doi.org/10.5194/hess-25-3397-2021

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, *7*, 53040–53065. https://doi.org/10.1109/ACCESS.2019.2912200

Sibanda, M., Mutanga, O., Chimonyo, V. G. P., Clulow, A. D., Shoko, C., Mazvimavi, D., Dube, T., & Mabhaudhi, T. (2021). Application of drone technologies in surface water resources monitoring and assessment: A systematic review of progress, challenges, and opportunities in the global south. *Drones*, *5*(3), 84. https://doi.org/10.3390/drones5030084

Smith, R. A., Schwarz, G. E., & Alexander, R. B. (1997). Regional interpretation of water-quality monitoring data. *Water Resources Research*, *33*(12), 2781–2798. https://doi.org/10.1029/97WR02171

Solanki, A., Agrawal, H., & Khare, K. (2015). Predictive Analysis of Water Quality Parameters using Deep Learning. *International Journal of Computer Applications*, *125*(9), 29–34. https://doi.org/10.5120/ijca2015905874

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, *9*(1), 62–72. https://doi.org/10.1177/089443939100900106

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed.). A Bradford Book.

Sprague, L. A., Oelsner, G. P., & Argue, D. M. (2017). Challenges with secondary use of multi-source water-quality data in the United States. *Water Research*, *110*, 252–261. https://doi.org/10.1016/j.watres.2016.12.024

Steefel, C. I. (2019). Reactive transport at the crossroads. *Reviews in Mineralogy and Geochemistry*, *85*(1), 1–26. https://doi.org/10.2138/rmg.2019.85.1

Steefel, C. I., Appelo, C. A. J., Arora, B., Jacques, D., Kalbacher, T., Kolditz, O., Lagneau, V., Lichtner, P. C., Mayer, K. U., Meeussen, J. C. L., Molins, S., Moulton, D., Shao, H., Šimůnek, J., Spycher, N., Yabusaki, S. B., & Yeh, G. T. (2015). Reactive transport codes for subsurface environmental simulation. *Computational Geosciences*, *19*(3), 445–478. https://doi.org/10.1007/s10596-014-9443-x

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, *338*(6106), 496–500. https://doi.org/10.1126/science.1227079

Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letters*, *14*(7), 073001. https://doi.org/10.1088/1748-9326/ab1b7d

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *ArXiv:1703.01365 [Cs]*. http://arxiv.org/abs/1703.01365

Tang, T., Strokal, M., van Vliet, M. T. H., Seuntjens, P., Burek, P., Kroeze, C., Langan, S., & Wada, Y. (2019). Bridging global, basin and local-scale water quality modeling towards enhancing water quality management worldwide. *Environmental Change Assessment*, *36*, 39–48. https://doi.org/10.1016/j.cosust.2018.10.004

Topp, S. N., Pavelsky, T. M., Jensen, D., Simard, M., & Ross, M. R. V. (2020). Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water*, *12*(1), 169. https://doi.org/10.3390/w12010169

Tsai, W.-P., Fang, K., Ji, X., Lawson, K., & Shen, C. (2020). Revealing causal controls of storage-streamflow relationships with a data-centric Bayesian framework combining machine learning and process-based modeling. *Frontiers in Water*, *2*, 40. https://doi.org/10.3389/frwa.2020.583000

Van Liew, M. W., Arnold, J. G., & Garbrecht, J. D. (2003). Hydrologic simulation on agricultural watersheds: Choosing between two models. *Transactions of the ASAE*, *46*(6), 1539–1551. https://doi.org/10.13031/2013.15643

van Vliet, M. T. H., Yearsley, J. R., Franssen, W. H. P., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2012). Coupled daily streamflow and water temperature modelling in large river basins. *Hydrology and Earth System Sciences*, *16*(11), 4303–4321. https://doi.org/10.5194/hess-16-4303-2012

Vanschoren, J. (2019). Meta-learning. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning: Methods, systems, challenges* (pp. 35–61). Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5_2

Varadharajan, C., Hendrix, V. C., Christianson, D. S., Burrus, M., Wong, C., Hubbard, S. S., & Agarwal, D. A. (2022). BASIN-3D: A brokering framework to integrate diverse environmental data. *Computers & Geosciences*, *159*, 105024. https://doi.org/10.1016/j.cageo.2021.105024

Varadharajan, C., Kumar, V., Willard, J., Zwart, J., Sadler, J., Weierbach, H., Perciano, T., Mueller, J., Hendrix, V., & Christianson, D. (2021). *Using machine learning to develop a predictive understanding of the impacts of extreme water cycle perturbations on river water quality* (AI4ESP1135, 1769795). https://doi.org/10.2172/1769795

Varanka, S., Hjort, J., & Luoto, M. (2015). Geomorphological factors predict water quality in boreal rivers. *Earth Surface Processes and Landforms*, *40*(15), 1989–1999. https://doi.org/10.1002/esp.3601

Vesselinov, V. V., Alexandrov, B. S., & O'Malley, D. (2018). Contaminant source identification using semi-supervised machine learning. *Journal of Contaminant Hydrology*, *212*, 134–142. https://doi.org/10.1016/j.jconhyd.2017.11.002

Wang, R., Kim, J.-H., & Li, M.-H. (2021). Predicting stream water quality under different urban development pattern scenarios with an interpretable machine learning approach. *Science of the Total Environment*, *761*, 144057. https://doi.org/10.1016/j.scitotenv.2020.144057

Weierbach, H., Lima, A. R., Willard, J. D., Hendrix, V. C., Christianson, D. S., Lubich, M., & Varadharajan, C. (2022). Stream temperature predictions for River Basin Management in the Pacific Northwest and mid-Atlantic Regions Using Machine Learning. *Water*, *14*(7), 1032. https://doi.org/10.3390/w14071032

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Westphal, K., Musolff, A., Graeber, D., & Borchardt, D. (2020). Controls of point and diffuse sources lowered riverine nutrient concentrations asynchronously, thereby warping molar N:P ratios. *Environmental Research Letters*, *15*(10), 104009. https://doi.org/10.1088/1748-9326/ab98b6

White, J. T. (2017). Forecast first: An argument for groundwater modeling in reverse. *Groundwater*, *55*(5), 660–664. https://doi.org/10.1111/gwat.12558

White, K. L., & Chaubey, I. (2007). Sensitivity analysis, calibration, and validations for a multisite and multivariable Swat model. *JAWRA Journal of the American Water Resources Association*, *41*(5), 1077–1089. https://doi.org/10.1111/j.1752-1688.2005.tb03786.x

Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., & Wade, A. J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal*, *54*(1), 101–123. https://doi.org/10.1623/hysj.54.1.101

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Willard, J., Read, J. S., Appling, A. P., Oliver, S. K., Jia, X., & Kumar, V. (2020). Predicting water temperature dynamics of unmonitored lakes with meta transfer learning. http://arxiv.org/abs/2011.05369

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ArXiv*:2003.04919 [Physics, Stat]. http://arxiv.org/abs/2003.04919

Wood, S. N., Goude, Y., & Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *64*(1), 139–155. https://doi.org/10.1111/rssc.12068

Xie, H., & Lian, Y. (2013). Uncertainty-based evaluation and comparison of SWAT and HSPF applications to the Illinois River basin. *Journal of Hydrology*, *481*, 119–131. https://doi.org/10.1016/j.jhydrol.2012.12.027

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., & Shen, C. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *Journal of Hydrology*, *603*, 127043. https://doi.org/10.1016/j.jhydrol.2021.127043

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *WIREs Water*, *8*(5), e1533. https://doi.org/10.1002/wat2.1533

Xu, Z., Serata, R., Wainwright, H., Denham, M., Molins, S., Gonzalez-Raymat, H., Lipnikov, K., Moulton, D., & Eddy-Dilek, C. (2021). Reactive transport modeling for supporting climate resilience at groundwater contamination sites. *Hydrology and Earth System Sciences Discussions*, 1–29. https://doi.org/10.5194/hess-2021-338

Yan, J., Gao, Y., Yu, Y., Xu, H., & Xu, Z. (2020). A prediction model based on deep belief network and least squares SVR applied to cross-section water quality. *Water*, *12*(7), 1929. https://doi.org/10.3390/w12071929

Yang, G., & Moyer, D. L. (2020). Estimation of nonlinear water-quality trends in high-frequency monitoring data. *Science of the Total Environment*, *715*, 136686. https://doi.org/10.1016/j.scitotenv.2020.136686

Yang, Y., & Chui, T. F. M. (2020). Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods. *Hydrology and Earth System Sciences Discussions*, 1–41. https://doi.org/10.5194/hess-2020-460

Yang, Y., & Chui, T. F. M. (2021). Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods. *Hydrology and Earth System Sciences*, *25*(11), 5839–5858. https://doi.org/10.5194/hess-25-5839-2021

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *ArXiv:2003.05689 [Cs, stat]*. http://arxiv.org/abs/2003.05689

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, *598*, 126266. https://doi.org/10.1016/j.jhydrol.2021.126266

Zhang, G. P. (2007). Avoiding pitfalls in neural network research. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *37*(1), 3–16. https://doi.org/10.1109/TSMCC.2006.876059

Zhang, Q., Blomquist, J. D., Moyer, D. L., & Chanat, J. G. (2019). Estimation bias in water-quality constituent concentrations and fluxes: A synthesis for Chesapeake Bay rivers and streams. *Frontiers in Ecology and Evolution*, *7*, 109. https://doi.org/10.3389/fevo.2019.00109

Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., & Li, L. (2021). From hydrometeorology to river water duality: Can a deep learning model predict dissolved oxygen at the continental scale? *Environmental Science & Technology*, *55*(4), 2357–2368. https://doi.org/10.1021/acs.est.0c06783

Zhu, S., & Piotrowski, A. P. (2020). River/stream water temperature forecasting using artificial intelligence models: A systematic review. *Acta Geophysica*, *68*(5), 1433–1442. https://doi.org/10.1007/s11600-020-00480-7

Zwart, J. A., Oliver, S. K., Watkins, W. D., Sadler, J. M., Appling, A. P., Corson-Dosch, H. R., Jia, X., Kumar, V., & Read, J. S. (2021). *Near-term forecasts of stream temperature using process-guided deep learning and data assimilation*. https://eartharxiv.org/repository/view/2590/

How to cite this article: Varadharajan, C., Appling, A. P., Arora, B., Christianson, D. S., Hendrix, V. C., Kumar, V., Lima, A. R., Müller, J., Oliver, S., Ombadi, M., Perciano, T., Sadler, J. M., Weierbach, H., Willard, J. D., Xu, Z., & Zwart, J. (2022). Can machine learning accelerate process understanding and decision-relevant predictions of river water quality? *Hydrological Processes*, *36*(4), e14565. https://doi.org/10.1002/hyp.14565