nature water

Review article

https://doi.org/10.1038/s44221-024-00202-z

Deep learning for water quality

Received: 13 April 2023

Accepted: 10 January 2024

Published online: 12 March 2024

Check for updates

Wei Zhi[1,2], Alison P. Appling[3], Heather E. Golden[4], Joel Podgorski[5] & Li Li[2]

Understanding and predicting the quality of inland waters are challenging, particularly in the context of intensifying climate extremes expected in the future. These challenges arise partly due to complex processes that regulate water quality, and arduous and expensive data collection that exacerbate the issue of data scarcity. Traditional process-based and statistical models often fall short in predicting water quality. In this Review, we posit that deep learning represents an underutilized yet promising approach that can unravel intricate structures and relationships in high-dimensional data. We demonstrate that deep learning methods can help address data scarcity by filling temporal and spatial gaps and aid in formulating and testing hypotheses via identifying influential drivers of water quality. This Review highlights the strengths and limitations of deep learning methods relative to traditional approaches, and underscores its potential as an emerging and indispensable approach in overcoming challenges and discovering new knowledge in water-quality sciences.

Artificial intelligence (AI) has been used for data processing since the 1930s and 1940s[1,2]. In World War Two, the Turing machine, an early form of AI, saved an estimated 20 million lives by decoding data encrypted by the German Enigma[3]. The term 'deep learning', however, was not coined until 1986[4], after the emergence of classic algorithms such as recurrent neural network (RNN) and convolutional neural network (CNN) in the 1970s (Boxes 1 and 2). Deep learning (DL), an AI method characterized by multiple hidden layers (≥2), has experienced a recent renaissance since 2006[5,6]. This renaissance has been catalysed by novel algorithms without the need for domain expertise and human supervision and the advent of graphical and tensor processing units (GPUs and TPUs). These advances have enabled automatic extraction of complex patterns and relationships[7], igniting an explosion of applications in almost every discipline. Earth and environmental sciences are no exception[8]. DL has been used for predicting flooding and sediments since the late 1990s[9], although its expanded use in hydrology is relatively recent (since 2016)[10,11]. The application of DL in water quality, however, has lagged behind[12,13] (Box 2).

Here we posit that DL presents promising opportunities for addressing water-quality challenges where process-based, statistical, and even other machine learning (ML) approaches have frequently fallen short, particularly because DL can predict water quality and fill data gaps by leveraging diverse, widely available data. In particular, DL can predict sparsely measured water-quality variables and detect patterns in highly complex relationships. Here we (1) describe the challenges in water-quality sciences that DL can help to resolve, (2) review opportunities for DL in water quality prediction, particularly in addressing data scarcity and in fostering new knowledge, (3) introduce emerging tools such as process-guided DL (PGDL), differentiable modelling (DM) and explainable DL (XDL) methods, and 4) offer a forward-looking perspective on the future of water-quality prediction.

This Review focuses on current literature and future directions of DL particularly on water-quality-related topics. Review papers on AI and ML applications in other topics abound, including, for example, a general DL introduction[7], DL in hydrology[9,14] and ecosystem science[15], and ML in marine science[16], environmental and water management[17,18], crop yield mapping[19], environmental science and engineering[20], inland water quantity, quality and ecology[21], decision-relevant prediction and management[13], and differentiable modelling[22], to name just a few. Interested readers are referred to these reviews on relevant topics.

[1]The National Key Laboratory of Water Disaster Prevention, Yangtze Institute for Conservation and Development, Key Laboratory of Hydrologic-Cycle and Hydrodynamic-System of Ministry of Water Resources, Hohai University, Nanjing, China. [2]Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, USA. [3]US Geological Survey, Reston, VA, USA. [4]Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH, USA. [5]Department of Water Resources and Drinking Water, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Dübendorf, Switzerland. e-mail: lili@engr.psu.edu

https://doi.org/10.1038/s44221-024-00202-z

BOX 1

Deep learning glossary

Artificial intelligence (AI) broadly describes machine intelligence that can simulate human intelligence, such as learning, reasoning and problem solving.

Machine learning (ML) is a subfield of AI that uses algorithms and statistical models to enable machines to learn from data and make predictions or decisions without being explicitly programmed.

Neural network (NN) is a type of ML algorithm inspired by the structure and function of biological neural networks in human brain. It includes neurons as the basic building blocks that are organized into input, output and hidden layers.

Deep learning (DL) is a subset of neural network with deeper networks, typically with multiple hidden layers (≥2).

DL techniques include recurrent neural network (RNN), convolutional neural network (CNN), autoencoder, long short-term memory (LSTM), deep belief network (DBN), gated recurrent unit (GRU), generative adversarial network (GAN) and transformer. Detailed information on their structure and function can be found in refs. 9,21. Other DL-related acronyms used in this work include graphical processing units (GPUs), tensor processing units (TPUs), stochastic gradient descent (SGD), process-guided deep learning (PGDL), differentiable modelling (DM), explainable deep learning (XDL), integrated gradients (IG), expected gradients (EG), Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME).

Although different DL algorithms share the common feature of having multiple hidden layers to automatically learn from raw data, CNN is well suited for spatial analysis tasks such as processing image data, whereas RNN, LSTM, GRU and transformer are more suitable for sequential tasks such as time-series prediction. DBN is useful for feature extraction, for example, to identify commonalities among water bodies or water-quality patterns. Autoencoder and GAN can produce realistic complex data such as images and parameter maps and can also automatically detect anomalies, for example, contamination events in a water supply network[142,143].

In a typical DL algorithm, raw input data are processed through multiple layers, each transforming data for automated extraction and learning of hierarchical, nonlinear and complex representations[7]. The advent of powerful computing resources, such as software to leverage GPUs and TPUs, has enabled the training of increasingly complex and deeper neural networks, boosting the breadth of DL applications. The advancement in structure also enhanced computational efficiency. For example, CNNs have utilized local connectivity, shared weights, pooling layers and deep architectures to reduce parameter numbers. RNNs use feedback connections and backpropagation through time to predict based on an entire sequence of steps, incorporating information about recent and cumulative events in the context of the timing and order of their occurrence. Although these techniques have been around for decades, their applications in water sciences have become prevalent only in recent years[144].

Long-standing challenges in water quality

Water quality has been degrading worldwide under the compound stresses of direct and indirect human influence[23], including increased pollution loads from urbanization and agricultural expansion[24], more frequent and prolonged duration of hypoxia in a warming climate[25], and persistent and widespread harmful algal blooms (HABs)[26]. Understanding and predicting water quality are therefore imperative but have faced major challenges. First, water chemistry is complex and encompasses many variables. For example, the US Geological Survey maintains an inventory of up to 24,898 water-quality variables across 17 categories (https://help.waterdata.usgs.gov/codes-and-parameters/parameters). In this Review, we cast a wide but incomplete net by searching the literature for DL modelling of common water-quality term (Supplementary Text). Their transport and fate are regulated by interacting, complex processes under distinct environmental conditions. These multiple layers of complexity make it challenging to measure, model, understand and predict water quality.

Challenges with data scarcity

Data collection and measurements are the foundation of scientific discovery. They enable the formulation of hypotheses and the development of conceptual and numerical models. Compared with streamflow, water-quality data, however, are often more sparse, inconsistent, and limited in time, space and frequency[27] (Fig. 1), partly owing to the complexity of water-quality variables. Common water-quality measurements include water temperature (WT), total suspended solids (TSS), dissolved oxygen (DO), biological and chemical oxygen demand, salinity, specific conductance, turbidity, sediments, clarity, alkalinity, chlorophyll *a* (chl *a*), carbon and nutrients in various forms (for example, dissolved organic carbon, nitrate (NO_3), total nitrogen (TN), total phosphorous (TP) and toxic metals (for example, arsenic, lead). Most water-quality variables still require manual and labour-intensive measurements using 'grab samples' and chemical analysis using large, complex analytical instruments, in contrast to hydrological data (for example, precipitation, streamflow, evapotranspiration and snow depth) that are often measured automatically[27]. Although sensors have been developed for hundreds of water-quality variables[28], their in situ deployment for routine, automatic measures are limited in location, duration and water-quality variables (for example, WT, DO, specific conductance, nutrients, dissolved organic carbon). Even for the most abundantly measured TSS, the global average is limited to 29 data counts per station, 1.1% of days with data and a record duration of 4.2 years (Fig. 1), compared with 12,066, 84% and 38 years for streamflow. Some sites do have long-term water-quality records, although they are a small fraction of the total (outliers in Supplementary Fig. 1).

In addition, data availability is highly heterogeneous: approximately 83% of global TSS data comes from 17% of the sampled rivers, predominantly in North America. Other variables have even lower coverages (Fig. 1b–d). An additional limitation of grab samples is that water-quality monitoring often fails to capture the full range of streamflow regimes (for example, transient peak flow), often leading to bias in modelling water quality. Furthermore, monitoring efforts are often patchy, tailored to address specific environmental concerns, such as a summer algal bloom event or metal pollution resulting from a mine-waste spill. These localized, short-term datasets may have limited applicability for broader assessments or long-term trend analyses. Note that most examples in the following section are US based due to the availability of openly accessible, long-term water-quality datasets; however, relevant DL studies in other global locations are discussed wherever possible.

Challenges with model prediction

Predicting water-quality dynamics remains a major challenge yet is essential for water management, risk mitigation and climate adaptation[29]. Linear statistical approaches, including those integrating mass balances, are meaningful screening tools to assess drivers (for example, climate, urbanization, agricultural expansion). They, however, are typically limited by the assumption of linear and/or stationary relationships between drivers, concentrations or loads of focal variables, and therefore cannot simulate changing dynamics and predict future

https://doi.org/10.1038/s44221-024-00202-z

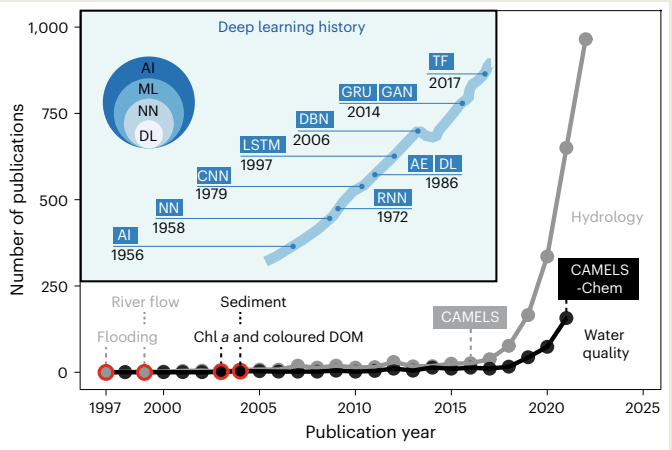
BOX 2

A brief history from AI to DL and beyond

The term AI was not coined until 1956 (inset in figure)[145], although the idea originated in the 1930s and 1940s when Alan Turing first published about computing machinery and intelligence[1,2]. The concept of neural network was first proposed in 1943[146]; the first trainable neural network was demonstrated in 1958[147]. Although DL approaches such as RNN emerged as early as 1972[148], the term 'deep learning' was not coined until 1986[4]. The approach has been revived in representational learning since 2006[5], as detailed by Schmidhuber[145].

DL has been used for prediction and knowledge discovery since the 1970s; shallow neural networks (for example, artificial neural network) have been used to predict water quality since the 1990s[149,150]. Yet DL application in water resources has gained momentum only in recent years. Early DL applications in hydrology (for example, flooding[151] and river flow[152]) and water quality[153,154] (for example, chl*a*, coloured dissolved organic matter and sediment) used the multilayer perceptron neural network, although one hidden layer was sometimes used to reduce training time and overfitting[155]. A period of quiescent, scattered publications followed until 2017 when the CAMELS database was published[156,157]. The CAMELS database inspired other datasets including Global Streamflow Indices and Metadata Archive (GSIM)[136] and CAMELS in individual countries, and the global community dataset Caravan (published in 2023 and thus not shown on the curve that extends to 2022)[158]. These datasets likely have facilitated DL application in hydrology, as indicated by the skyrocketing rise (grey line in the figure), although some popular algorithms (for example, LSTM, RNN, CNN) have been around for two or three decades (inset in figure).

DL publications on water quality have lagged by a few years, with one-fifth and one-quarter of the publications compared with those in hydrology by 2021 and 2022, respectively, although part of the differences may arise from the community size differences. The advent of water-quality databases such as GEMStat[159], Global River Chemistry (GLORICH) database[160], Surface Water Chemistry (SWatCh) database[161], Global River Water Quality Archive (GRQA)[137] and CAMELS-Chem[162] may similarly accelerate DL application in water quality.



History of DL in hydrology and water quality. The number of peer-reviewed deep learning publications (up to the end of 2022) was retrieved from the Web of Science (Supplementary Information). The grey–red circles indicate the first DL publications in hydrology on flooding[151,155] (1997, 1999) and river flow[152] (1999). The black–red circles indicate the first DL publications on water quality for chl*a*[154,163] and coloured dissolved organic matter (DOM)[55] in 2003 and 2004, and sediments[153] (2004). CAMELS[162] is a benchmark dataset for large-scale hydrology in the United States. A suite of similar datasets was developed in other countries (for example, Chile, Great Britain, Brazil, Australia, France, Switzerland), leading to the development of global Caravan[158]. Similarly, the stream chemistry database CAMELS-Chem[162] integrates atmospheric and stream water chemistry data with hydrometeorological data from CAMELS, which is often needed for DL models. Inset: a brief history of concepts and milestones, where advances are chronologically ordered but not to scale in the vertical or horizontal directions. The concentric circles indicate the subset relationship between AI, ML, NN and DL.

water-quality conditions. Other non-ML statistical models similarly have limited flexibility and adaptability to changing conditions. For example, LoadEST (Load Estimator)[30] and WRTDS (Weighted Regressions on Time, Discharge, and Season)[31] are primarily based on the relationships between concentration, discharge and time. These estimates can be compromised when these relationships vary and depend on unmodelled factors[32]. ML has gained popularity due to its ability to analyse and extract patterns from large and complex datasets without relying on explicit physical or chemical equations. However, traditional ML models often require manual engineering for feature extraction from input data and struggle to capture long-term temporal dependencies in scarce data. This is particularly the case for water-quality data. For example, the global average TSS record duration is 4.2 years per site, far from sufficient for capturing long-term trends.

Process-based models are another model category for water-quality predictions. These models typically solve ordinary or partial differential equations based on mass-balance principles of water and chemical variables and explicitly simulate underlying processes that govern water-quality dynamics. One of their major strengths is to provide insights into mechanisms of water-quality dynamics[33], as they are guided by physics and chemistry principles. Process-based models, however, suffer from several major limitations. In most cases, we lack a comprehensive, mechanism-based understanding such that processes may not be accurately and adequately represented in the models[34]. Process-based models also require detailed data on a myriad of processes and properties[35], including above- and below-ground characteristics, water flow, and biogeochemical processes, which are time-consuming and expensive to collect. Process-based models are also computationally expensive, particularly when simulating at large spatio-temporal scales and resolutions. They are also limited because extrapolation from one variable to another often requires different process representation and calibration data, and therefore model re-development or re-calibration, even within the same watershed[33]. This is challenging because water-quality concerns vary by space and time.

Deep learning approaches

Strengths of DL approaches

DL approaches can provide high predictive accuracy[36] and have the potential to address long-standing challenges facing traditional statistical and process-based models. DL models are flexible, adaptable, integrative, scalable and speedy. They are flexible, in that they can learn complex relationships from raw data without requiring a careful feature engineering of inputs and a detailed understanding of underlying processes[7], making them useful in deciphering high-dimensional

https://doi.org/10.1038/s44221-024-00202-z

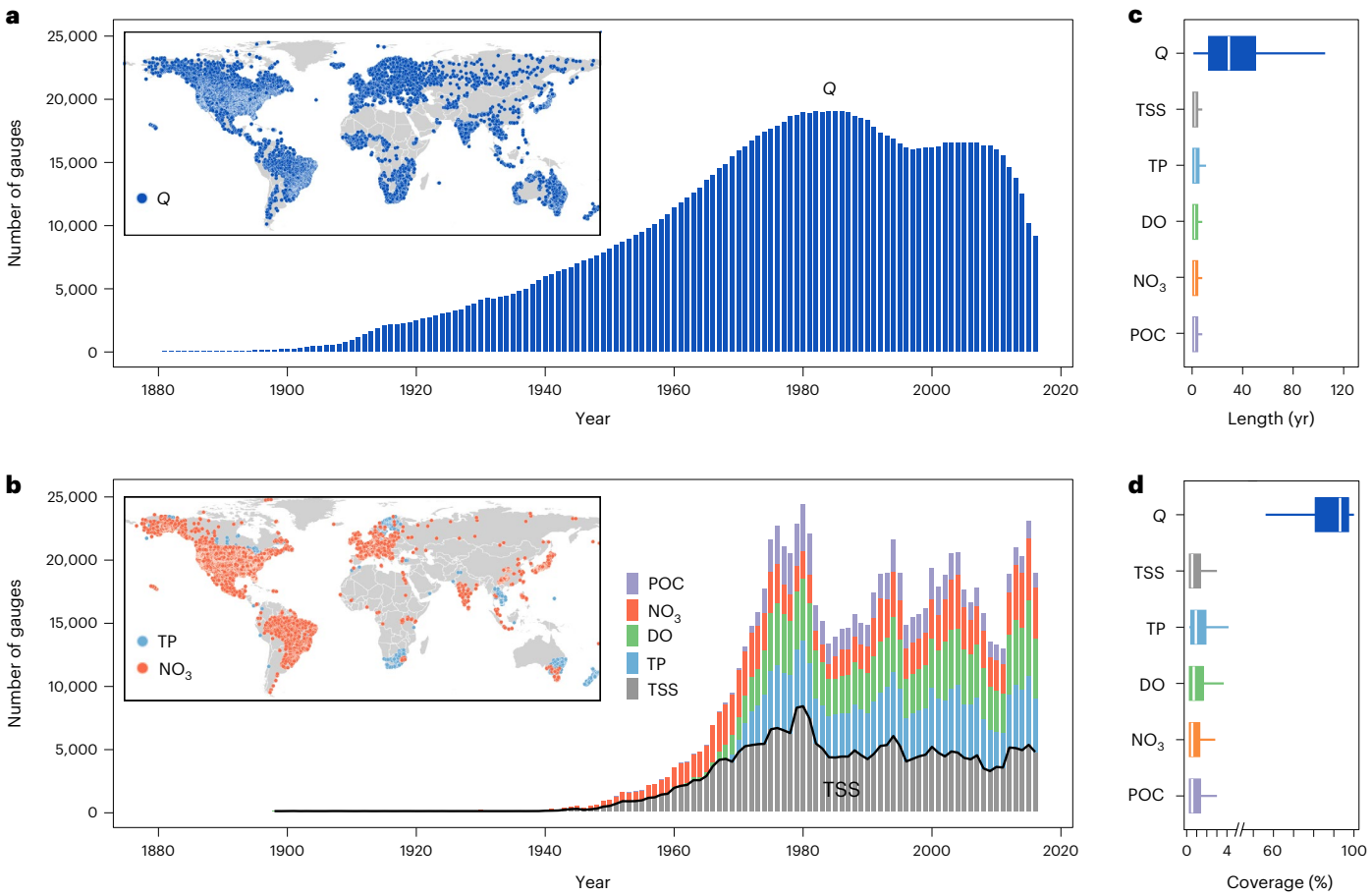


Fig. 1 | Gauges through ages and across a few representative water-quality variables. a,**b**, Temporal trends of the number of global gauges reporting at least one data point for streamflow (*Q*) (**a**) and water quality (**b**), with data from the Global Streamflow Indices and Metadata Archive (GSIM)[136] and the Global River Water Quality Archive (GRQA)[137], respectively. The inset maps show global gauge locations. **c**,**d**, The outliers in Supplementary Fig. 1. The 25%, 50% (middle line) and 75% percentiles of data length (**c**) and temporal coverage (**d**). Streamflow *Q* has a total of 374 million data points from 30,959 sites, whereas TSS has a total of ~2 million data points from 68,592 sites. Other variables include TP (1.9 million data points from 44,943 sites), DO (1.2 million data points from 48,066 sites), NO_3 (1.2 million from 44,551 sites) and particulate organic carbon (POC; 0.62 million data points from 22,877 sites). The length (yr, **c**) is the number of years that have data points. The coverage (%, **d**) is the temporal coverage percentage with data points in days. Streamflow (*Q*) gauging started in the United States in the 1880s and increased steadily until 1960s[138], when gauges began to expand across Europe and other continents (**a**). A decrease in streamflow gauges since 2015 may indicate withering investment[139], or a latency in data mobilization: it takes time for new observations to become publicly available. The first gauge for TSS, the most abundant water-quality variable, was established in the United States in 1898[137] (**b**), almost 20 years after the first streamflow gauge. The spike in observations in the 1970s probably arose from the substantial water infrastructure investments from the Clean Water Act[140]. Noticeable declines followed around 1980 and 1995 are possibly due to funding cuts[141]. Observations in the GLORICH database extend only until 2011 (see breakdown in Supplementary Fig. 2), leading to another decline after 2010 (**b**).

environmental data with poorly understood mechanisms. Second, they are adaptable, in that they can learn from new data without requiring substantial modifications of model structure. For example, the same model can often be used when new water-quality observations or management-relevant datasets become available. They are integrative, as they can extract hidden patterns and nonlinear representations from diverse data sources[8], such as sensor data, satellite images and grab chemistry data that vary in availability and spatio-temporal coverage. They are scalable as they are designed to learn directly from data that already embed information on spatial and temporal scales[9], which can reduce the need for model parameters at particular scales that are important yet often unavailable via measurements. Such parameters include, for example, local hyporheic exchange and solute transformation rates. DL models are also speedy, as they take advantage of both hardware advances and optimization algorithms that are designed to efficiently traverse high-dimensional parameter space and converge quickly, enabling the exploration of many environmental scenarios and prediction of many parameters across broad spatial and temporal extents.

In some instances, DL can also represent physical processes in climate and geoscience models, where these processes might be inadequately understood and coarsely modelled[8,22]. For example, deep neural networks (DNNs) have been applied to represent turbulent processes in ocean models[37] and atmospheric subgrid processes in climate models[38] to minimize the prohibitive cost of running high-resolution physical models. The saved computational resources can then be reallocated to enhance simulations either by increasing ensemble sizes or by improving the model resolution[39]. These advantages and features, as well as the use of problem-specific DL architectures such as CNN for spatial analysis and RNN for time-series tasks (Box 1), are well suited to modelling the complex and spatiotemporally dynamic nature of water-quality conditions.

Limitations of DL approaches

Despite gaining tremendous momentum, DL models have limitations. In addition to requiring significant computational resources such as GPUs and TPUs, DL models require enormous datasets to

https://doi.org/10.1038/s44221-024-00202-z

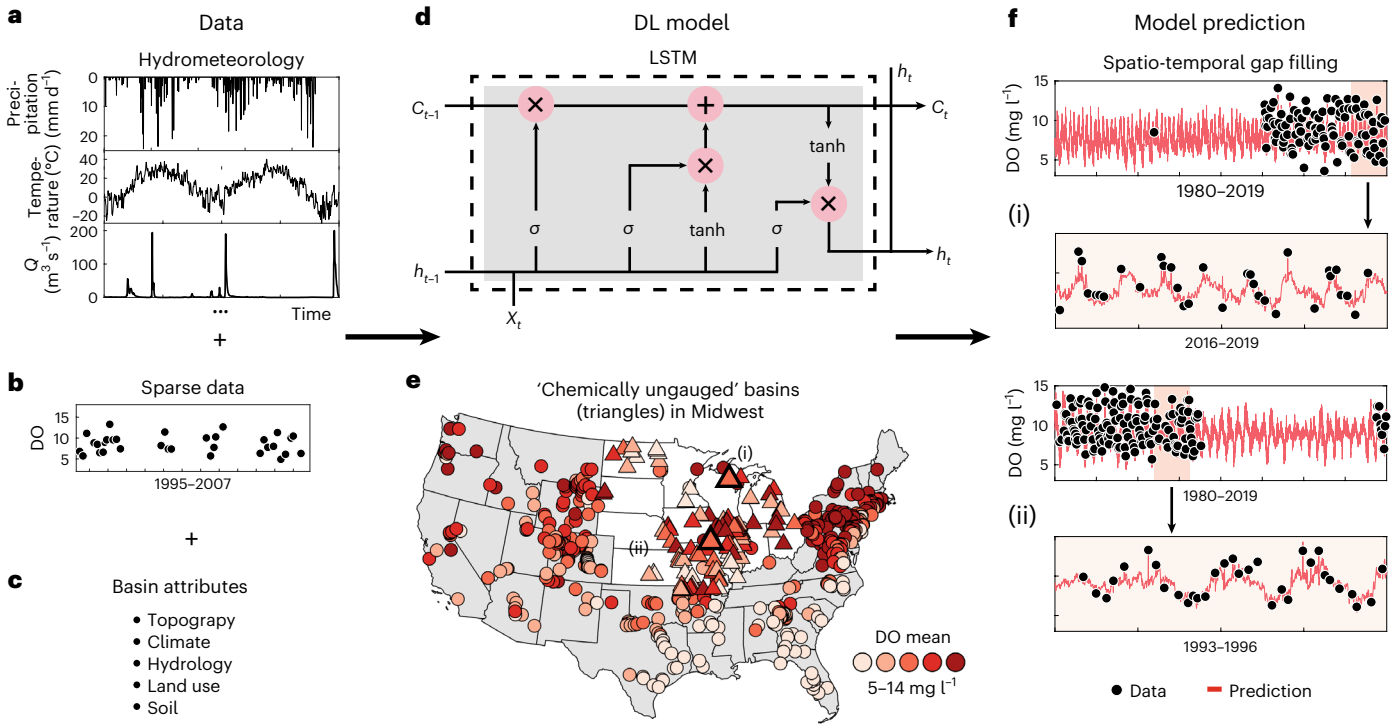


Fig. 2 | The use of a DL model for spatio-temporal water-quality gap filling. **a**–**e**, A continental-scale LSTM model (**d**) was trained with hydrometeorological data (**a**), sparse DO data (**b**) and constant basin attributes (**c**) from 480 rivers to predict DO dynamics in 100 'chemically ungauged basins' (blank white region with triangles in US map; **e**), where data were excluded during training. **f**, The DL model robustly reproduced the long-term (1980–2019) DO trends and seasonal variations (zoom in) in these 'data excluded' rivers, indicating its potential in predicting DO in chemically ungauged basins. In **d**, X_t, C_t and h_t represent the input, cell state and hidden state at the current timestep t, respectively. The symbols σ and tanh refer to sigmoid and tanh functions, respectively. Pink circles denoted by × and + correspond to point-by-point multiplication and addition operations, respectively. Panels adapted with permission from: **a**,**b**,**d**, ref. 12, American Chemical Society; **e**,**f**, ref. 40, Springer Nature Ltd.

train effectively. Without sufficient data, they are prone to overfitting[9]. That is, they become too closely tailored to the training data and fail to aptly generalize to new conditions. In other words, DL models are only as good as their data; if they have not seen enough data, for example, under extreme conditions, they cannot learn to extract the input–output relationship under these conditions and are not better than traditional models. In hydrology, the availability of large benchmark datasets (for example, Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS)) has evolved concurrently with fast-growing DL applications (Box 2). DL applications in water quality have grown comparatively slowly, potentially indicating data limitations as a bottleneck (Fig. 1 and Box 2). DL models are additionally criticized as being 'black boxes' and lack interpretability and generalizability, such that it is challenging to understand mechanisms and extrapolate beyond training data. These limitations have triggered advances in PGDL, as discussed in later sections.

Deep learning for data-scarcity challenges

The challenges of data scarcity cannot be resolved overnight. Data collection requires investments in physical and human resources and technological innovations, including the development of new sensors that can automatically measure variables under more frequent and intensifying extreme conditions. Yet the need to understand and predict water quality is urgent as we face pressing water-quality issues under changing climate conditions and human stresses. With ample data, DL models can predict water quality at times and locations without observations (spatial and temporal data filling) and help discover new information through model and data interrogation. In fact, recent work has leveraged publicly available data, including satellite imagery and hydrometeorology data, to predict water quality in surface and subsurface waters with scarce data.

Spatial data filling in chemically ungauged basins

Prediction from well-monitored to ungauged and unmonitored locations has been a long-standing challenge. DL models have recently shown promises in making prediction for chemically ungauged basins. Water-quality data in monitored locations have been used to build models together with hydrometeorological data, remote-sensing data or spatial features such as basin characteristics, and then extrapolate to ungauged rivers. For example, a continental-scale long short-term memory (LSTM) model trained with DO data from 480 US rivers made robust predictions in 100 rivers where data were purposely excluded from the training dataset to resemble ungauged rivers[40] (Fig. 2). LSTMs trained with process-based model predictions and WT observations from 145 well-monitored lakes achieved better performance than a pure process-based model of lake temperature when transferred to 1,882 less-monitored lakes in the Midwest United States[41]. A deep gated recurrent unit (GRU) model combined satellite images with relatively limited in situ measurements (that is, 1,260 pairs of water-clarity data from 399 lakes) to infer water clarity in 16,475 global lakes with little or no data[42]. Spatially explicit DL models have filled spatial gaps using the effects of unmonitored reaches on their neighbours to infer water quality at all reaches. Spatial relationships among stream reaches have been represented through graph convolution on an adjacency matrix based on stream distance[43] or travel time[44] for temperature prediction, or even detailed process-based routing within a DL context for streamflow prediction[45]. Such relationships can be made more nuanced by learning more specifics of reaches, such as those with reservoirs and those without[46] or for learned clusters of physically similar reaches[47].

Temporal data filling

DL models have been used to predict time series of water quality by incorporating spatial features, temporal correlations and nonlinearity without prior assumptions. Such capabilities underscore its potential for effectively filling temporal data gaps. For example, using time series of intensively measured hydrometeorological data, sparse DO data and static watershed characteristics as inputs, a trained LSTM model predicted daily DO in hundreds of US rivers[12,40] (Fig. 2). A regional multi-site LSTM model reproduced and gap-filled daily NO_3 measurements at 42 monitored stream reaches in Iowa[48] with improved performance (Supplementary Table 1). Furthermore, a modified LSTM model[49] combined hydrometeorological data and physical properties of lakes (for example, coordinates, elevation, surface area) and predicted daily WT from 1980 to 2020 in 185,549 US lakes[50]. LSTM models have also used spatial information from adjacent groundwater wells to enhance the accuracy of temporal gap filling for specific conductance, especially for large decadal gaps in the Columbia River[51]. Another study demonstrated that an LSTM model outperformed other ML models (for example, support vector machine, single-layer perceptron) in predicting daily TSS concentrations in a Malaysian river[52]. A hybrid encoder–decoder bidirectional LSTM model showed higher accuracy than ML (extreme gradient boosting) and standalone DL methods in predicting daily sediment loads in the Godavari River Basin in India[53].

Predicting data-scarce variables from data-rich surrogates

Water chemistry sensors have been increasingly deployed in recent years at temporal resolutions as fine as minutes; they, however, are limited to a handful of variables (for example, turbidity, specific conductance, pH, WT, DO and NO_3) with scant spatio-temporal coverage. Most water-quality variables are manually measured at low frequencies (for example, monthly, quarterly). Traditional remote sensing for water quality has primarily focused on optically active variables such as chl *a*[54], coloured dissolved organic matter DOM (DOM)[55] and water clarity[56] in large water bodies (oceans and large lakes and rivers); remote-sensing data often have insufficient spatial resolution for small rivers and streams.

Many water-quality variables, however, are intrinsically linked by shared transport dynamics, redox conditions and biogeochemical processes including, for example, soil respiration and nutrient transformation[57]. These relationships among variables have long been acknowledged and leveraged by surrogate regression models for sediments[58], pesticides[59] and nutrients[60], among others. Data-rich variables can therefore provide information about data-poor variables. DL approaches are now beginning to explore this opportunity. For example, DL models have recently been used to estimate nutrients that are non-optically active, based on their correlations with optically active variables that were estimated via remote sensing. A backpropagation neural network model trained with limited measurements and satellite-retrieved sea surface salinity and remote-sensing reflectance successfully estimated NO_3 and phosphate concentrations[61]. Another DL model estimated TN and TP based on their correlations with chl *a* and remote-sensing reflectance, and further reconstructed spatio-temporal patterns of nutrients from 1984 to 2020[62]. These estimated variables can further predict other variables, such as dissolved carbon, that are less frequently measured but are essential for water quality, and CO_2–climate feedbacks[63]. LSTM- and GRU-based models have been used to estimate TP and heavy metal concentrations (that is, copper, zinc) in urban sewer networks from commonly measured variables (for example, temperature, pH, conductivity)[64,65].

Predicting groundwater quality from catchment properties

Earth's subsurface, or the critical zone from soils to parent bedrock[66], governs the storage and flow of groundwater, biogeochemical reactions and chemical transport from groundwater to surface waters, and, therefore, surface water quality[67]. The subsurface, however, is not as readily accessible (for example, via boreholes or geophysics), such that below-ground data are even more scarce[68,69].

Subsurface properties and functions are generally not as temporally variable as those at the surface. Deep CNN-based DL models, therefore, are often used to extract spatial patterns. DL models have been shown to outperform traditional calibration approaches in estimating subsurface parameters, because they can directly infer parameters from observations and capture the high nonlinearity with fewer realizations[70,71]. For example, a deep CNN-based model recently used two-dimensional land-surface data, including digital elevation maps and remote-sensing images, to construct three-dimensional subsurface structures in an Australian desert landscape[72]. The model revealed complex relationships between surface and subsurface features that are often obscured by traditional methods such as sequential Gaussian. The DL model automated a low-cost method to generate a three-dimensional subsurface structure that inherits the probability structure of a real two-dimensional surface image. Another DNN model used widely available time-series streamflow data to estimate permeability, an essential property that determines flow rates that are arduous to measure directly[73], outperforming traditional ensemble smoother methods.

Non-DL algorithms have also been used to estimate spatial variations of mean groundwater quality. For example, random forest and generalized boosted regression models have been increasingly applied to predict groundwater contaminants from local to global scales[74–77]. An ML model trained with groundwater chemistry (for example, alkalinity, Ca, Mg, turbidity) was used to detect anomalous methane in groundwater in areas of shale gas production across the United States[78]. ML models have also been used to fill global gaps of groundwater contaminants including fluoride[75] and arsenic[74] (Fig. 3). In fact, groundwater chemistry may be better predicted by ML methods that can explicitly link environmental variables to spatial variability in groundwater chemistry but require fewer data. In many groundwater cases, predicting spatial variation may be more important than temporal trends, because prediction maps can help identify areas of low groundwater quality. Future opportunities lie in developing DL models that can train well with scarce data. For instance, some groundwater solutes may originate from the same geological setting. In these cases, transfer models could be trained on a larger dataset and then used to perform learning on less-measured water-quality variables.

Deep learning for robust predictions

A common concern about DL is its limited generalizability, that is, capability to extrapolate beyond the training data[9]. Unlike traditional process-based models, DL models usually rely solely on patterns in training data, which may be scarce especially under climate extremes such as fires, floods and droughts[57]. Advances to improve model performance with existing knowledge primarily reside in two directions: process-guided deep learning (PGDL) and differential modelling (DM). As shown in examples below, most existing applications of PGDL and DM are more in the realm of improving model prediction and parameter estimation. These approaches are expected to reveal process-based understanding and new knowledge but have yet to demonstrate such capabilities. PGDL and DM applications in water quality have been limited so far; we therefore also draw examples from hydrology to illustrate their potential use.

Process-guided deep learning

PGDL seeks to encode domain knowledge within otherwise domain-agnostic model architectures and training algorithms[79]. By doing so, the PGDL model leverages well-established process knowledge and discourages violation of known principles, which also helps earn stakeholder trust such that stakeholders use model outputs more readily. An advantage of PGDL over DL is the improved accuracy and reliability beyond training conditions[80]. PGDL can also improve the

https://doi.org/10.1038/s44221-024-00202-z

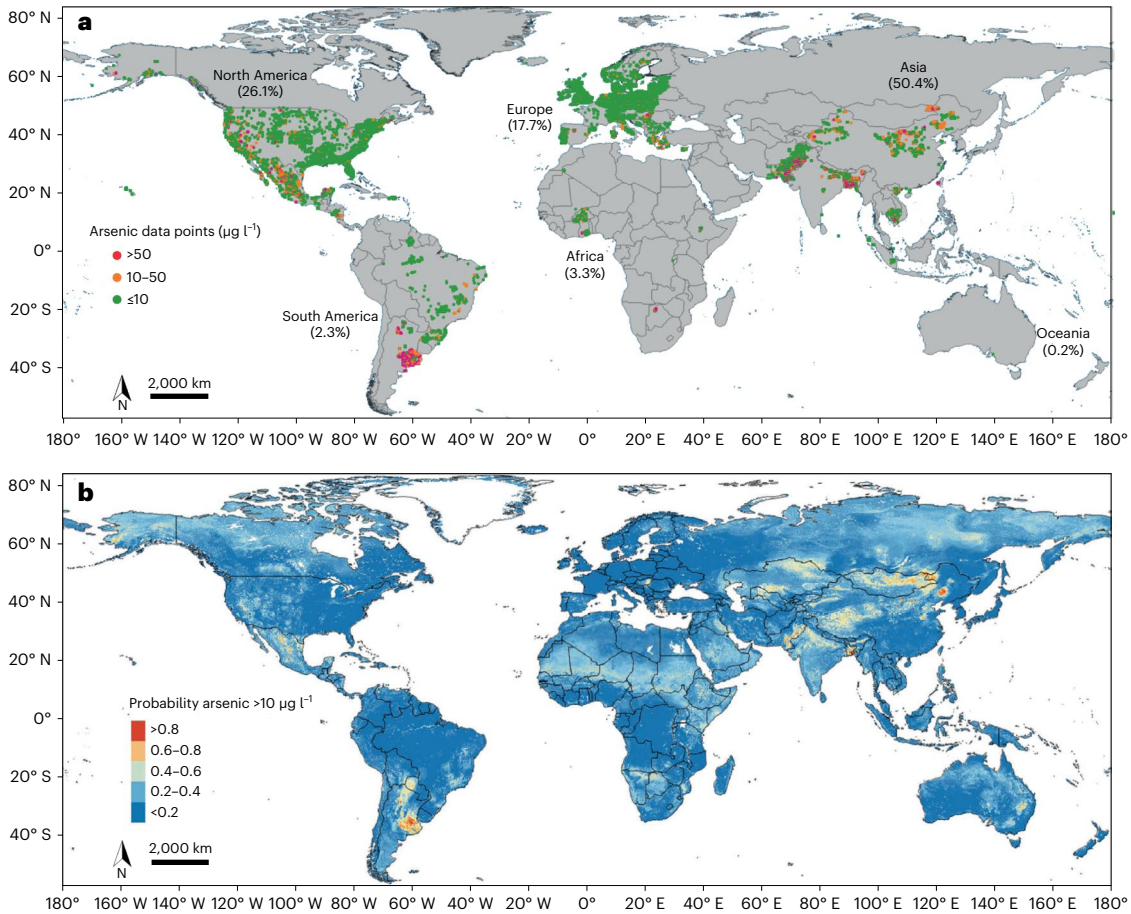


Fig. 3 | Global maps of groundwater arsenic. a, Map of 55,000 groundwater arsenic concentrations used as input to ML models. **b**, A global map of probability of arsenic in groundwater exceeding 10 μg l^{-1} (ref. 74). Some groundwater chemistry does not vary as much as in surface water, such that ML models without time dependence my be sufficient for generating spatial prediction maps. This example shows the potential of data-driven methods in estimating scarce groundwater chemistry. Figure adapted with permission from ref. 74, AAAS.

physical realism of DL predictions where data are limited, noisy or incomplete. One PGDL approach is to use output from process-based models (not necessarily calibrated) as additional training data[81], which augments the availability of data. PGDL can also be achieved by using physically meaningful loss function terms, such as a penalty for the lack of mass or energy conservation[82,83], or by adding asymmetric activation functions[84] to enforce constraints imposed by underlying processes, such as flow, transport and energy balance. Loss terms and constraints that explicitly encode hydrological or biogeochemical knowledge may look much like process-based models, with the advantage that the PGDL elements are written in a DL language that provides automatic differentiation, that is, calculation of gradients of outputs relative to all model variables. PGDL thus overlaps with the emerging field of DM, which closely interweaves process-based model equations and neural networks in a single differentiable language[85]. As an example, a hybrid physics-guided RNN model for lake temperature[86] incorporated energy conservation and density–depth relations into the loss function as penalty terms. The model was pretrained using simulated energy budgets from the physics-based General Lake Model to initialize the network structure and fill in scarce data. The model performed better and can project to warmer and colder conditions beyond training data. Similarly, an LSTM model pretrained with an energy budget formulation and WT predictions from the General Lake Model performed robustly when extended to 68 lakes outside of the training conditions[82].

Some PGDL methods additionally utilize multi-task learning, where DL models are trained to simultaneously predict related variables, such as streamflow and stream temperature, to encourage the learning of process-relevant information shared between variables[87]. For instance, a physics-informed neural network for subsurface solute transport[88] incorporated Darcy's law and advection–dispersion equations in a DL model and trained it together with hydraulic conductivity, hydraulic head and solute concentrations. The approach predicted concentrations of a synthetic solute that better matched a synthetic dataset than the standard DL model, especially when the training data were sparse. The model accuracy further improved when multiple variables were jointly inverted. Compared with single-task models, jointly predicted stream temperature and flow may have better performance, especially when hyperparameters are carefully tuned[87].

Differentiable modelling

DM aims to integrate process-based equations with DL models to simultaneously advance process representations, parameter estimation and predictive accuracy[22]. DM encodes existing knowledge and neural networks in an automatically differentiable programming language to reap the advantages of the physical underpinnings of process-based models and the learning capabilities of DL. DM includes physically meaningful parameters and equations that can be inspected and/or manipulated. DM can additionally approach the predictive accuracy of purely data-driven DL, suggesting that the DL components of a DM model can learn relationships that are encoded by process-based components. Recent analyses showed that DM with a physical model as the backbone can outperform pure DL, yielding more accurate regional extrapolation of streamflow with respect to daily metrics and decadal trends[80]. Similarly,

https://doi.org/10.1038/s44221-024-00202-z

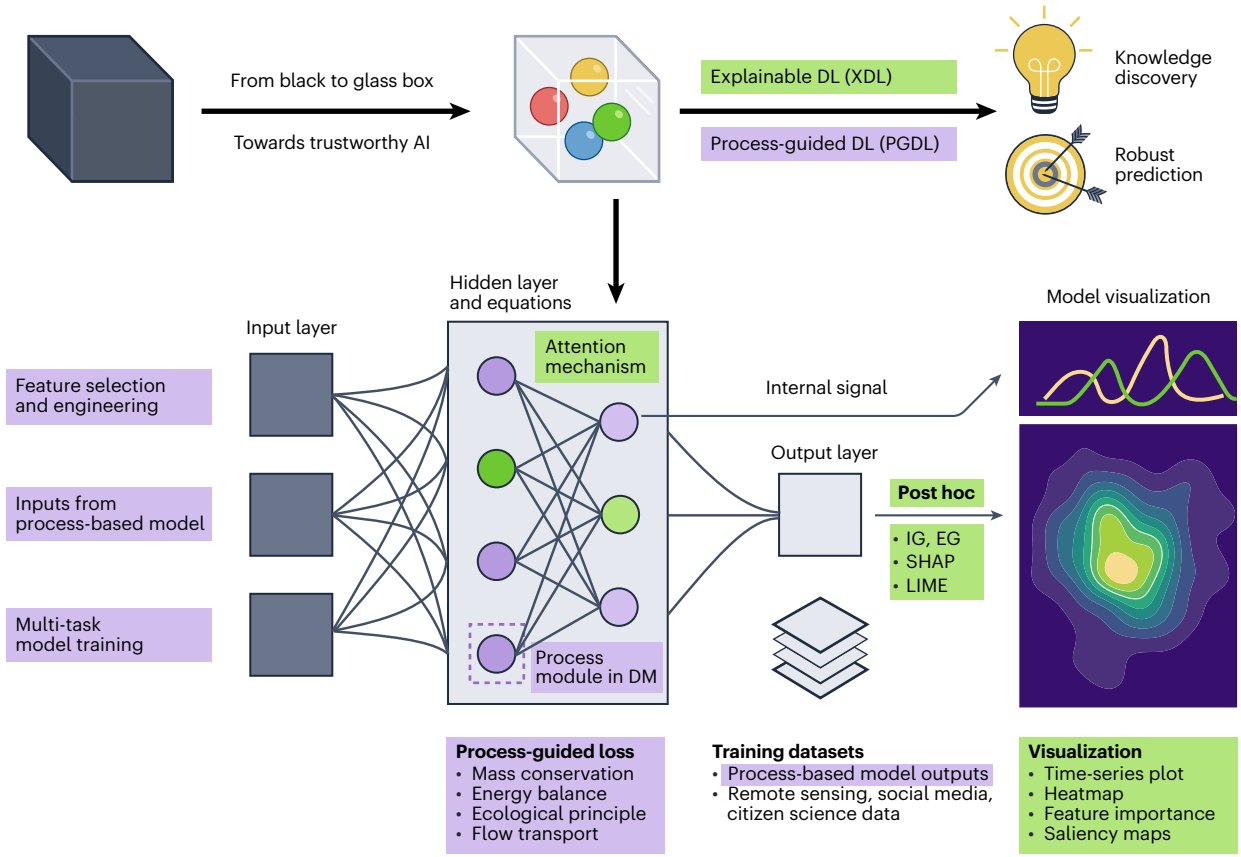


Fig. 4 | A conceptual diagram showing the ideas and approaches from black boxes to glass boxes towards robust model performance and knowledge discovery. The efforts include using PGDL (purple), DM and XDL (green) along with revealing visualizations. Domain knowledge can be integrated into DL at various stages, such as selecting important features, pretraining DL models using outputs from process-based models and multi-task learning. Alternatively, process-guided loss functions or differentiable process-based modules (for example, dashed box) can be used to incorporate mass conservation, energy balance, flow transport or other process knowledge to enhance model performance. Knowledge discovery can emerge from accurate predictions themselves, from inspecting variables and learned parameters within the model itself (for example, internal signals), and from XDL. XDL includes common post hoc methods such as integrated gradients (IG), expected gradients (EG), SHAP and LIME, and model-specific methods such as attention mechanisms.

embedding the hydrologic model EXP-HYDRO within an RNN structure and augmenting it with neural network layers accurately captured snow water equivalent and transferred streamflow prediction across different rivers[89]. A recent work introduced neural networks to substitute ordinary differential equations for representing hydrologic processes[90]. The results showed comparable performance to DL methods, surpassing a conceptual hydrologic model in streamflow prediction for 569 US rivers while retaining the interpretability of the conceptual model. Furthermore, a process-based model integrated an advective dispersion equation with a river network graph and predicted stream WT more accurately in data-sparse situations[44].

Deep learning for knowledge discovery

DL approaches have been criticized as being 'black boxes'[9]: the algorithms find the optimal combination of layers and weight functions to fit data without offering insights into mechanisms. Such a black-box approach does not reveal its inner workings and new knowledge of processes. With increasing awareness of this limitation, the pursuit of methods to judge the trustworthiness of DL approaches is growing, aiming to turn black boxes into transparent glass boxes for interpretability and knowledge discovery (Fig. 4). The toolbox of such techniques is growing[91]. Explainable deep learning (XDL) approaches aim to illuminate the 'black box' by evaluating model 'reasoning', interpreting model decisions, and extracting patterns and drivers (Fig. 4). XDL includes model-agnostic and model-specific approaches that identify and rank important features, relationships and mechanisms that contribute to model predictions[92]. Model-agnostic concepts include integrated gradients[93], expected gradients[94], Shapley additive explanations (SHAP)[95] and surrogate models such as local interpretable model-agnostic explanations (LIME)[96]. They do not require a specific model structure and therefore can provide comparable outputs for different models. Model-specific approaches include attention mechanisms, saliency maps and decision trees, and can tailor explanations for specific models (for example, transformers, CNNs and tree-based algorithms, respectively). These techniques elucidate 'behaviours' of deep learning[97,98] and support hypothesis generation. Hypothesis testing is essential for falsifying assumptions and theories and uncovering potentially overlooked patterns and correlations[99]. Consequently, this process fosters knowledge discovery, enhances process-based understanding, and facilitates more interpretable prediction and informed decision-making.

XDL has been used mostly in understanding temporal trends, spatial patterns, and predominant drivers of streamflow, water temperature (WT), and a limited number of water-quality variables. For example, XDL has been used to understand the spatial relationships of stream temperature and the seasonal importance of streamflow versus wind and air pressure in controlling saltwater intrusion into the Delaware River[100]. Saliency maps have been used to highlight the most important regions of an input image for predicting streamflow[98], suggesting that global sea surface temperatures influence river flows via atmospheric convection and teleconnections. Another study used SHAP values and identified WT, DO and TP as the most influential drivers of riverine chl *a*, a widely used indicator of harmful algae blooms

(HABs)[101]. A DNN model predicted a variety of water-quality variables, from which SHAP values identified the most influential factors[102]. In addition, hybrid models integrating existing knowledge and DL are promising in potentially advancing both prediction accuracy and process-based understanding. For example, a hybrid model for lake phosphorous combined an RNN model with ecological principles (for example, power scaling)[103]. The model predicted short-term and long-term variations in observed phosphorous with high accuracy, outperforming the process-based model and RNN alone. The model identified lake level and thermocline depth as the most important drivers of phosphorous loads and revealed a decade-long downwards trend as contributing to the long and slow change in phosphorous loads. The model further suggested that including an additional temperature component can improve the process-based model.

The future of deep learning in water quality

As DL becomes increasingly applied, tested and improved in old and new regimes, DL will probably become increasingly trustworthy for predicting future water quality under various management, policy, climate and socioeconomic scenarios. As Earth's climate evolves, climate extremes such as floods, droughts, cyclones and fires will become more frequent and severe. Such extremes often alter concentrations and loads of sediments and solutes by orders of magnitude[57,104]. During extreme wildfires, for example, sediment-loaded water often overwhelms water treatment plants[105]; during droughts, DO often drops to critically low levels and endangers aquatic ecosystems[106]. Water-quality hindcast and near-term forecasts, therefore, will be essential for designing water infrastructure and making real-time decisions on water and ecosystem management. In addition, the growing challenges of water-quality management will make trustworthy forecasts and scenario projections increasingly valuable. DL can be potentially leveraged not only for extreme events forecasting but also for general management such as identifying pollution sources[107], optimizing monitoring networks[108] and management decisions[109], and automatically monitoring water quality[54,110]. Such predictions are critical for adapting to climate changes and mitigating the impact of extreme events.

Approaches for hindcasts and forecasts will continue to face challenges of data scarcity and incomplete process understanding, although the approaches described above can begin to ameliorate these challenges. Furthermore, new developments in alternative methods such as Bayesian modelling[111], evolutionary algorithms[112] and transfer learning[113], as well as their hybrid use with DL models, could be leveraged for improved prediction. For example, a DL-guided evolutionary algorithm was trained to use sensor data to identify contamination sources and improve computational efficiency and model performance[114]. A transfer-learning-based LSTM model captured the long-term dependencies among time series and leveraged knowledge learned from complete datasets, improving imputation accuracy by 15–25% for DO concentration[115].

Existing work on water quality, whether using traditional DL, XDL, PGDL or DM, has only scratched the surface of our capacity to learn from DL models. Most work has been limited to a few variables such as WT and DO that are largely influenced by meteorological conditions, sediment and phosphorus that are primarily driven by discharge regimes, as well as optically active variables such as chl *a* and coloured DOM that can be directly inferred from the spectral signatures of satellite images. For now, every addition to the literature is valuable in developing our sense of what is possible and how to make the best possible use of DL, not only in practical uses such as forecasting extreme events but also in further developing theories and insights that drive water-quality dynamics.

Prediction for extreme events and climate scenarios

Water management under extreme conditions traditionally relies on human expertise (for example, subjective detection thresholds) and ensemble models for extreme weather prediction. However, models such as LSTM have shown promise in forecasts under extreme conditions with lead times of up to days[116]. As extreme events intensify and alter water quality, traditional process-based models may be limited by our understanding of water-quality theory under extreme conditions[57]. Existing data, if measured under extreme conditions (a big 'if' for water quality), may already contain valuable information that surpasses our current understanding[117]. Such hidden knowledge in data can be leveraged in DL models to forecast water quality under extreme conditions. As an example, HABNet, a model that combines CNN and LSTM, has discriminated between HAB and non-HAB events using remote-sensing data, outperforming historical methods[54]. An integrated PGDL and data assimilation approach forecasted daily WT up to 7 days in advance with accuracy and quantified uncertainties[109], enabling water management decisions such as reservoir water release when WT rose above a fish tolerance threshold.

Such existing work is only the tip of the iceberg. We anticipate that DL-based forecasting will expand beyond algae blooms and WT. The bottleneck of such forecasting is still sufficient data under extreme conditions[118]. To train well, DL models have to see sufficient input to output response to figure out trends and patterns. Extreme conditions challenge data collection, because physical conditions during, for example, floods, often prevent manual data collection but also knock out sensors used for automatic data collection. Extreme events, although predicted to occur more frequently, will still occur less frequently such that the temporal window for monitoring is fleeting. Advances in technology for robust and automatic measurements are essential in both gushing waters and in close-to-zero-flow dry riverbeds[119].

Predicting the future and projecting hypothetical scenarios into the future demand more than a capability for hindcasts. Data-driven models of all kinds (statistical, ML broadly and DL specifically) may predict accurately on training data but fail spectacularly under new input conditions. Generalizability demands that new conditions we wish to project are represented in the training datasets, which is often not the case. XDL can help evaluate the physical realism of DL predictions under diverse conditions, and PGDL can encourage DL models to encode physically realistic relationships. Rigorous tests of new DL models are needed with respect to generalizability. These include conducting spatio-temporal extrapolation tests in ungauged basin and future (lead) forecast, as well as in benchmark tests against other established methods or models. These tests will offer the capability to represent complex processes and project to new scenarios with more confidence and transparency, supporting decision-makers in anticipating and responding to water-quality challenges. Progress has been made with encouraging preliminary findings[82,100]; DL generalizability, however, should not be taken for granted.

Diversifying data sources to combat data scarcity

The challenges of data scarcity will continue, because data collection requires investment, human resources and innovation. Data scarcity can be ameliorated with expanded use of traditional DL, PGDL and DM in conjunction with observations of surrogates and other biogeochemically related predictors. However, data 'generated' by DL-based approaches should be used with caution. Another approach is to leverage an even wider diversity of data sources. For example, hydrology data are much more available than water-quality data. Remote-sensing data, social media data and citizen science data have become widely available. Social media posts, including text, pictures and videos, have been mined for flood-water level estimation[120], flood assessment[121] and water-quality classification[122]. Citizen science has also become increasingly useful in hydrological and water-quality research[123–125]. Cost-effective crowdsourced monitoring can additionally engage the public, thereby enhancing the long-term sustainability of monitoring networks[123]. For example, community-based monitoring provides water-quality data (for example, pH, WT, electrical conductivity) in

Chile[124] and Australia[126] (https://www.waterwatch.org.au). Citizen science data, however, may be challenging for water-quality variables, as measurements of most solutes require expensive technology and changes in solute concentrations are often invisible. Yet they can potentially provide 'complementary' information on environmental conditions or human behaviour for DL models to learn, infer and forecast water quality.

Seeking new knowledge

Earth's subsurface governs water storage, transport and the generation of water-quality variables via biogeochemical reactions[127], therefore regulating the chemistry of subsurface source water that eventually enters rivers and lakes[57]. In fact, a significant portion of surface waters derive from soil water and groundwater[128]. Surface water chemistry therefore reflects water flow paths and its interactions with soils, rocks, microbes and roots that mobilize solutes and sediments[129]. Existing theories and empirical relationships abound depicting how physical, chemical and biological processes mobilize solids and solutes[40,43,130]. Simultaneous use of both process-based and DL models, whether independently or within a coupled framework, can inspire new hypotheses about mechanisms and drivers of water-quality dynamics (Fig. 4).

XDL tools can potentially be used to compare PGDL and non-PGDL models and reveal what the models can learn differently when we ask and/or enable them to better conform to a physics-based reality. Interrogating XDL of multi-task models may reveal influential predictors or latent variables the models learn to produce and share among multiple variables. The learned relationships can further generate hypotheses. Theories and process-based equations combined with DL modules in the DM framework may also enable rapid calibration and comparison of competing process representations for hypothesis testing[22]. In addition, model interrogation with different types of input can reveal influential drivers. For example, assessment of model performance with different inputs have revealed temperature as the predominant driver of daily DO in US rivers[40]. Robust model training offers consistent 'data' output with filled gaps, which can enable extraction of temporal trends and spatial patterns. For example, a multi-task LSTM model trained on WT and DO data in about 800 rivers revealed that rivers warm up most rapidly in urban rivers and lose oxygen most rapidly in agricultural rivers, and that they lose oxygen faster than oceans[131]. These approaches could open doors for knowledge discovery.

The opportunities for combined XDL, PGDL and DM to inform knowledge exist, although they have yet to realize their full potential. Current DL research has focused more on approaches and model performance than on knowledge discovery. The emerging data (whether original observation or model filled) and knowledge from these approaches can have far-reaching impacts, not only on water-quality prediction but also on broad understanding of processes that shape global biogeochemical cycles of carbon, nutrients and other elements, and climate feedbacks[132].

In summary, while transparent, interpretable process-based and statistical models will remain important for predicting water quality, DL models can potentially overcome long-standing data limitations and predictive challenges inherent to these traditional approaches. DL approaches, however, are only as robust as the quality and quantity of the available data. Data and observations are the bedrock of all scientific discoveries[133,134]. Without intercepted messages, Turing could not have decoded Enigma[135]. Similar to children learning to speak, DL models must 'see' or 'hear' enough data to decipher hidden patterns and laws. Using diverse data sources, including remote sensing, social media content, citizen science data, surrogate water-quality measurements and process-based model outputs, can potentially improve our ability to leverage DL for understanding and predicting water quality. Despite these additional sources, data availability will probably remain the bottleneck of DL applications in water quality. Paradoxically, with the 'right' amount of data, DL models can help predict water quality in time and space, filling data gaps and reconstructing long-term data. A potential future direction is to integrate DL and process-based models (for example, PGDL, DM), harnessing their individual merits for improved prediction, transparency, and knowledge discovery.

We predict that the emergent power of DL approaches for improving global water quality will be realized through: (1) collating publicly available spatial and temporal data and exploring their relationships with water-quality variables for spatio-temporal prediction; (2) bringing new tools and fresh eyes to discover hidden patterns, processes and relationships that regulate water-quality dynamics; and 3) predicting future and unmonitored water-quality conditions to explore options for managing and mitigating water-quality impairments under climate extremes, and broadly in a rapidly changing world. The outcome could have far-reaching impacts not only in water-quality fields but also broadly in understanding and predicting the future of global cycles of carbon, nutrients, other elements and beyond.

Data availability

Streamflow data (Fig. 1a) from the Global Streamflow Indices and Metadata Archive (GSIM) were compiled from repositories at https://doi.org/10.1594/PANGAEA.887477 and https://doi.org/10.1594/PANGAEA.887470. Water-quality data (Fig. 1b) from the Global River Water Quality Archive (GRQA) were downloaded from https://doi.org/10.5281/zenodo.7056647.

References

1. Turing, A. M. Computing machinery and intelligence. *Mind* **59**, 433–460 (1950).
2. Turing, A. M. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **s2-42**, 230–265 (1937).
3. Muggleton, S. Alan Turing and the development of artificial intelligence. *AI Commun.* **27**, 3–10 (2014).
4. Dechter, R. Learning while searching in constraint-satisfaction problems. In *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 5, 178–183 (AAAI, 1986).
5. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
6. Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
7. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
8. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
9. Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **54**, 8558–8593 (2018).
10. Shen, C. P. et al. HESS opinions: incubating deep-learning-powered hydrologic science advances as a community. *Hydrol. Earth Syst. Sci.* **22**, 5639–5656 (2018).
11. Xu, T. & Liang, F. Machine learning for hydrologic sciences: an introductory overview. *Wiley Interdiscip. Rev. Water* **8**, e1533 (2021).
12. Zhi, W. et al. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? *Environ. Sci. Technol.* **55**, 2357–2368 (2021).
13. Varadharajan, C. et al. Can machine learning accelerate process understanding and decision-relevant predictions of river water quality? *Hydrol. Process.* https://doi.org/10.1002/hyp.14565 (2022).
14. Tripathy, K. P. & Mishra, A. K. Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions. *J. Hydrol.* https://doi.org/10.1016/j.jhydrol.2023.130458 (2023).
15. Perry, G. L. W., Seidl, R., Bellvé, A. M. & Rammer, W. An outlook for deep learning in ecosystem science. *Ecosystems* **25**, 1700–1718 (2022).

16. Song, T. et al. A review of artificial intelligence in marine science. *Front. Earth Sci.* https://doi.org/10.3389/feart.2023.1090185 (2023).
17. Sun, A. Y. & Scanlon, B. R. How can big data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environ. Res. Lett.* **14**, 073001 (2019).
18. Zhu, J.-J., Yang, M. & Ren, Z. J. Machine learning in environmental research: common pitfalls and best practices. *Environ. Sci. Technol.* **57**, 17671–17689 (2023).
19. van Klompenburg, T., Kassahun, A. & Catal, C. Crop yield prediction using machine learning: a systematic literature review. *Comput. Electron. Agric.* **177**, 105709 (2020).
20. Zhong, S. et al. Machine learning: new ideas and tools in environmental science and engineering. *Environ. Sci. Technol.* https://doi.org/10.1021/acs.est.1c01339 (2021).
21. Appling, A. P., Oliver, S. K., Read, J. S., Sadler, J. M. & Zwart, J. A. in *Encyclopedia of Inland Waters* 2nd edn (eds Mehner, T. & Tockner, K.) 585–606 (Elsevier, 2022).
22. Shen, C. et al. Differentiable modelling to unify machine learning and physical models for geosciences. *Nat. Rev. Earth Environ.* **4**, 552–567 (2023).
23. Diamond, J. S. et al. Hypoxia is common in temperate headwaters and driven by hydrological extremes. *Ecol. Indic.* **147**, 109987 (2023).
24. Creed, I. F. et al. Enhancing protection for vulnerable waters. *Nat. Geosci.* **10**, 809–815 (2017).
25. Nazari-Sharabian, M., Ahmad, S. & Karakouzian, M. Climate change and eutrophication: a short review. *Eng. Technol. Appl. Sci. Res.* **8**, 3668 (2018).
26. Paerl, H. W., Otten, T. G. & Kudela, R. Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum. *Environ. Sci. Technol.* **52**, 5519–5529 (2018).
27. McCabe, M. F. et al. The future of Earth observation in hydrology. *Hydrol. Earth Syst. Sci.* **21**, 3879–3914 (2017).
28. Abbott, B. W. et al. Using multi-tracer inference to move beyond single-catchment ecohydrology. *Earth Sci. Rev.* **160**, 19–42 (2016).
29. Milly, P. C. D. et al. Stationarity is dead: whither water management? *Science* **319**, 573–574 (2008).
30. Runkel, R. L., Crawford, C. G. & Cohn, T. A. *Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers* Report no. 4-A5 (USGS, 2004).
31. Hirsch, R. M., Moyer, D. L. & Archfield, S. A. Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay River inputs. *J. Am. Water Resour. Assoc.* **46**, 857–880 (2010).
32. Zhang, Q., Blomquist, J. D., Moyer, D. L. & Chanat, J. G. Estimation bias in water-quality constituent concentrations and fluxes: a synthesis for Chesapeake Bay rivers and streams. *Front. Ecol. Evol.* https://doi.org/10.3389/fevo.2019.00109 (2019).
33. Zhi, W. et al. Distinct source water chemistry shapes contrasting concentration–discharge patterns. *Water Resour. Res.* **55**, 4233–4251 (2019).
34. Archfield, S. A. et al. Accelerating advances in continental domain hydrologic modeling. *Water Resour. Res.* **51**, 10078–10091 (2015).
35. Zhi, W. et al. BioRT-Flux-PIHM v1.0: a watershed biogeochemical reactive transport model. *Geosci. Model Dev.* **15**, 19 (2022).
36. Rahmani, F. et al. Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. *Environ. Res. Lett.* https://doi.org/10.1088/1748-9326/abd501 (2021).
37. Bolton, T. & Zanna, L. Applications of deep learning to ocean data inference and subgrid parameterization. *J. Adv. Model. Earth Syst.* **11**, 376–399 (2019).
38. Rasp, S., Pritchard, M. S. & Gentine, P. Deep learning to represent subgrid processes in climate models. *Proc. Natl Acad. Sci. USA* **115**, 9684–9689 (2018).
39. Irrgang, C. et al. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **3**, 667–674 (2021).
40. Zhi, W., Ouyang, W., Shen, C. & Li, L. Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers. *Nat. Water* **1**, 249–260 (2023).
41. Willard, J. D. et al. Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. *Water Resour. Res.* **57**, e2021WR029579 (2021).
42. He, Y. et al. Water clarity mapping of global lakes using a novel hybrid deep-learning-based recurrent model with Landsat OLI images. *Water Res.* **215**, 118241 (2022).
43. Jia, X. et al. Physics-guided recurrent graph model for predicting flow and temperature in river networks. In *Proc. 2021 SIAM International Conference on Data Mining (SDM)* 612–620 (SIAM, 2021); https://doi.org/10.1137/1.9781611976700.69
44. Bao, T. et al. Partial differential equation driven dynamic graph networks for predicting stream water temperature. In *2021 IEEE International Conference on Data Mining (ICDM)* 11–20 (IEEE, 2021); https://doi.org/10.1109/ICDM51629.2021.00011
45. Bindas, T. et al. Improving river routing using a differentiable Muskingum-Cunge model and physics-informed machine learning. *Water Resour. Res.* **60**, e2023WR035337 (2024).
46. Chen, S. et al. Heterogeneous stream-reservoir graph networks with data assimilation. In *2021 IEEE International Conference on Data Mining (ICDM)* 1024–1029 (IEEE, 2021).
47. Chen, S., Zwart, J. A. & Jia, X. Physics-guided graph meta learning for predicting water temperature and streamflow in stream networks. In *Proc. 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 2752–2761 (Association for Computing Machinery, 2022).
48. Saha, G. K., Rahmani, F., Shen, C., Li, L. & Cibin, R. A deep learning-based novel approach to generate continuous daily stream nitrate concentration for nitrate data-sparse watersheds. *Sci. Total Environ.* **878**, 162930 (2023).
49. Kratzert, F. et al. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **23**, 5089–5110 (2019).
50. Willard, J. D., Read, J. S., Topp, S., Hansen, G. J. A. & Kumar, V. Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020). *Limnol. Oceanogr. Lett.* https://doi.org/10.1002/lol2.10249 (2022).
51. Ren, H., Cromwell, E., Kravitz, B. & Chen, X. Using long short-term memory models to fill data gaps in hydrological monitoring networks. *Hydrol. Earth Syst. Sci.* **26**, 1727–1743 (2022).
52. Latif, S. D. et al. Sediment load prediction in Johor River: deep learning versus machine learning models. *Appl. Water Sci.* https://doi.org/10.1007/s13201-023-01874-w (2023).
53. Jamei, M. et al. Designing a decomposition-based multi-phase pre-processing strategy coupled with EDBi-LSTM deep learning approach for sediment load forecasting. *Ecol. Indic.* **153**, 110478 (2023).
54. Hill, P. R., Kumar, A., Temimi, M. & Bull, D. R. HABNet: machine learning, remote sensing-based detection of harmful algal blooms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 3229–3239 (2020).
55. D'Alimonte, D., Zibordi, G. & Berthon, J. F. Determination of CDOM and NPPM absorption coefficient spectra from coastal water remote sensing reflectance. *IEEE Trans. Geosci. Remote Sens.* **42**, 1770–1777 (2004).

56. Zhang, Y. et al. Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods. *GIsci. Remote Sens.* **59**, 1367–1383 (2022).
57. Li, L. et al. River water quality shaped by land–river connectivity in a changing climate. *Nat. Clim. Change* https://doi.org/10.1038/s41558-023-01923-x (2024).
58. Rasmussen, P. P., Gray, J. R., Glysson, G. D. & Ziegler, A. C. *Guidelines and Procedures for Computing Time-Series Suspended-Sediment Concentrations and Loads from In-Stream Turbidity-Sensor and Streamflow* Report No. 3-C4 (USGS, 2009).
59. Covert, S. A., Bunch, A. R., Crawford, C. G. & Oelsner, G. P. *Comparison of Surrogate Models to Estimate Pesticide Concentrations at Six U.S. Geological Survey National Water Quality Network Sites during Water Years 2013–18* Report No. 2022-5109 (USGS, 2023).
60. Schilling, K. E., Kim, S.-W. & Jones, C. S. Use of water quality surrogates to estimate total phosphorus concentrations in Iowa rivers. *J. Hydrol. Reg. Stud.* **12**, 111–121 (2017).
61. Wang, D. et al. Satellite retrieval of surface water nutrients in the coastal regions of the East China Sea. *Remote Sens.* **10**, 1896 (2018).
62. Guo, H. et al. Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. *ISPRS J. Photogramm. Remote Sens.* **183**, 451–469 (2022).
63. Kerins, D. & Li, L. High dissolved carbon concentration in arid rocky mountain streams. *Environ. Sci. Technol.* **57**, 4656–4667 (2023).
64. Zhang, Y. et al. A hybrid model combining mode decomposition and deep learning algorithms for detecting TP in urban sewer networks. *Appl. Energy* **333**, 120600 (2023).
65. Jiang, Y., Li, C., Song, H. & Wang, W. Deep learning model based on urban multi-source data for predicting heavy metals (Cu, Zn, Ni, Cr) in industrial sewer networks. *J. Hazard. Mater.* https://doi.org/10.1016/j.jhazmat.2022.128732 (2022).
66. Li, L. et al. Expanding the role of reactive transport models in critical zone processes. *Earth Sci. Rev.* **165**, 280–301 (2017).
67. Li, L. et al. Toward catchment hydro-biogeochemical theories. *WIREs Water* **8**, e1495 (2021).
68. Bergen, K. J., Johnson, P. A., de Hoop, M. V. & Beroza, G. C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **363**, eaau0323 (2019).
69. Kolbe, T. et al. Stratification of reactivity determines nitrate removal in groundwater. *Proc. Natl Acad. Sci. USA* **116**, 2494–2499 (2019).
70. Sun, A. Y. Discovering state-parameter mappings in subsurface models using generative adversarial networks. *Geophys. Res. Lett.* https://doi.org/10.1029/2018gl080404 (2018).
71. Jiang, P., Shuai, P., Sun, A., Mudunuru, M. K. & Chen, X. Knowledge-informed deep learning for hydrological model calibration: an application to Coal Creek Watershed in Colorado. *Hydrol. Earth Syst. Sci. Discuss.* **2022**, 1–31 (2022).
72. Jiang, Z. et al. Sub3DNet1.0: a deep-learning model for regional-scale 3D subsurface structure mapping. *Geosci. Model Dev.* **14**, 3421–3435 (2021).
73. Cromwell, E. et al. Estimating watershed subsurface permeability from stream discharge data using deep neural networks. *Front. Earth Sci.* https://doi.org/10.3389/feart.2021.613011 (2021).
74. Podgorski, J. & Berg, M. Global threat of arsenic in groundwater. *Science* **368**, 845–850 (2020).
75. Podgorski, J. & Berg, M. Global analysis and prediction of fluoride in groundwater. *Nat. Commun.* https://doi.org/10.1038/s41467-022-31940-x (2022).
76. Ransom, K. M., Nolan, B. T., Stackelberg, P. E., Belitz, K. & Fram, M. S. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Sci. Total Environ.* **807**, 151065 (2022).
77. Nolan, B. T., Green, C. T., Juckem, P. F., Liao, L. & Reddy, J. E. Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA. *J. Hydrol.* **559**, 428–441 (2018).
78. Wen, T., Liu, M., Woda, J., Zheng, G. & Brantley, S. L. Detecting anomalous methane in groundwater within hydrocarbon production areas across the United States. *Water Res.* **200**, 117236 (2021).
79. Willard, J., Jia, X., Xu, S., Steinbach, M. & Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.* **55**, 1–37 (2023).
80. Feng, D., Beck, H., Lawson, K. & Shen, C. The suitability of differentiable, learnable hydrologic models for ungauged regions and climate change impact assessment. *Hydrol. Earth Syst. Sci. Discuss.* **2022**, 1–28 (2022).
81. Sun, A. Y., Yoon, H., Shih, C.-Y. & Zhong, Z. Applications of physics-informed scientific machine learning in subsurface science: A survey. In *Knowledge Guided Machine Learning* (eds Karpatne, A. et al.) 111–132 (Chapman and Hall/CRC, 2022).
82. Read, J. S. et al. Process-guided deep learning predictions of lake water temperature. *Water Resour. Res.* **55**, 9173–9190 (2019).
83. Beucler, T., Rasp, S., Pritchard, M. & Gentine, P. Achieving conservation of energy in neural network emulators for climate modeling. Preprint at https://arxiv.org/abs/1906.06622 (2019).
84. Daw, A. et al. Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling. In *Proc. 2020 SIAM International Conference on Data Mining (SDM)* 532–540 (SIAM, 2020).
85. Shen, C. et al. Differentiable modelling to unify machine learning and physical models for geosciences. *Nat. Rev. Earth Environ.* https://doi.org/10.1038/s43017-023-00450-9 (2023).
86. Jia, X. et al. Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles. In *Proceedings of the 2019 SIAM International Conference on Data Mining* 558–566 (SIAM, 2019).
87. Sadler, J. M. et al. Multi-task deep learning of daily streamflow and water temperature. *Water Resour. Res.* **58**, e2021WR030138 (2022).
88. He, Q., Barajas-Solano, D., Tartakovsky, G. & Tartakovsky, A. M. Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Adv. Water Res.* **141**, 103610 (2020).
89. Jiang, S., Zheng, Y. & Solomatine, D. Improving AI system awareness of geoscience knowledge: symbiotic integration of physical approaches and deep learning. *Geophys. Res. Lett.* https://doi.org/10.1029/2020gl088229 (2020).
90. Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C. & Fenicia, F. Improving hydrologic models for predictions and process understanding using neural ODEs. *Hydrol. Earth Syst. Sci.* **26**, 5085–5102 (2022).
91. Molnar, C. *Interpretable Machine Learning* (Lulu.com, 2020).
92. Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J. & Muller, K.-R. Explaining deep neural networks and beyond: a review of methods and applications. *Proc. IEEE* **109**, 247–278 (2021).
93. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning* 3319–3328 (ICML, 2017).
94. Erion, G. et al. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nat. Mach. Intell.* **3**, 620–631 (2021).
95. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 1–10 (2017).
96. Ribeiro, M. T., Singh, S. & Guestrin, C. " Why should i trust you?" Explaining the predictions of any classifier. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144 (ACM, 2016).

97. Xie, W. et al. Interpretable framework of physics-guided neural network with attention mechanism: simulating paddy field water temperature variations. *Water Resour. Res.* **58**, e2021WR030493 (2022).
98. Liu, Y., Duffy, K., Dy, J. G. & Ganguly, A. R. Explainable deep learning for insights in El Niño and river flows. *Nat. Commun.* https://doi.org/10.1038/s41467-023-35968-5 (2023).
99. Sadayappan, K., Kerins, D., Shen, C. & Li, L. Nitrate concentrations predominantly driven by human, climate, and soil properties in US rivers. *Water Res.* **226**, 119295 (2022).
100. Topp, S. N. et al. Stream temperature prediction in a shifting environment: explaining the influence of deep learning architecture. *Water Resour. Res.* https://doi.org/10.1029/2022wr033880 (2023).
101. Lee, D. et al. Integrated explainable deep learning prediction of harmful algal blooms. *Technol. Forecast. Soc. Change* **185**, 122046 (2022).
102. Zheng, H., Liu, Y., Wan, W., Zhao, J. & Xie, G. Large-scale prediction of stream water quality using an interpretable deep learning approach. *J. Environ. Manage.* **331**, 117309 (2023).
103. Hanson, P. C. et al. Predicting lake surface water phosphorus dynamics using process-guided machine learning. *Ecol. Modell.* **430**, 109136 (2020).
104. Carpenter, S. R., Booth, E. G. & Kucharik, C. J. Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnol. Oceanogr.* **63**, 1221–1233 (2018).
105. Robinne, F.-N. et al. Scientists' warning on extreme wildfire risks to water supply. *Hydrol. Process.* **35**, e14086 (2021).
106. Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M. & Wade, A. J. A review of the potential impacts of climate change on surface water quality. *Hydrol. Sci. J.* **54**, 101–123 (2009).
107. Wang, P. et al. Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. *Sci. Total Environ.* **693**, 133440 (2019).
108. Kontos, Y. N., Kassandros, T., Katsifarakis, K. L. & Karatzas, K. Deep Learning Modeling of Groundwater Pollution Sources. In *International Conference on Engineering Applications of Neural Networks* 165–177 (Springer, 2021).
109. Zwart, J. A. et al. Near-term forecasts of stream temperature using deep learning and data assimilation in support of management decisions. *J. Am. Water Res. Assoc.* **59**, 317–337 (2023).
110. van Lieshout, C., van Oeveren, K., van Emmerik, T. & Postma, E. Automated river plastic monitoring using deep learning and cameras. *Earth Space Sci.* **7**, e2019EA000960 (2020).
111. Beria, H., Larsen, J. R., Michelon, A., Ceperley, N. C. & Schaefli, B. HydroMix v1.0: a new Bayesian mixing framework for attributing uncertain hydrological sources. *Geosci. Model Dev.* **13**, 2433–2450 (2020).
112. Tang, Y., Reed, P. & Wagener, T. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrol. Earth Syst. Sci.* **10**, 289–307 (2006).
113. Iman, M., Arabnia, H. R. & Rasheed, K. A review of deep transfer learning and recent advancements. *Technologies* **11**, 40 (2023).
114. Qian, K., Jiang, J., Ding, Y. & Yang, S.-H. DLGEA: a deep learning guided evolutionary algorithm for water contamination source identification. *Neural Comput. Appl.* **33**, 11889–11903 (2021).
115. Chen, Z. et al. A transfer learning-based LSTM strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system. *J. Hydrol.* **602**, 126573 (2021).
116. Xiang, Z., Yan, J. & Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **56**, e2019WR025326 (2020).
117. Nearing, G. S. et al. What role does hydrological science play in the age of machine learning? *Water Resour. Res.* https://doi.org/10.1029/2020WR028091 (2021).
118. Guo, D. et al. A data-based predictive model for spatiotemporal variability in stream water quality. *Hydrol. Earth Syst. Sci.* **24**, 827–847 (2020).
119. Zimmer, M. A. et al. Zero or not? Causes and consequences of zero-flow stream gage readings. *WIREs Water* **7**, e1436 (2020).
120. Chaudhary, P., D'Aronco, S., Moy de Vitry, M., Leitão, J. P. & Wegner, J. D. Flood-water level estimation from social media images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* https://doi.org/10.5194/isprs-annals-IV-2-W5-5-2019 (2019).
121. Kanth, A. K., Chitra, P. & Sowmya, G. G. Deep learning-based assessment of flood severity using social media streams. *Stoch. Environ. Res. Risk Assess.* **36**, 473–493 (2022).
122. Hanif, M., Khawar, A., Tahir, M. A. & Rafi, M. Deep learning based framework for classification of water quality in social media data. In *Proc. MediaEval 2021 Workshop (*MediaEval 2021).
123. Njue, N. et al. Citizen science in hydrological monitoring and ecosystem services management: state of the art and future prospects. *Sci. Total Environ.* **693**, 133531 (2019).
124. Yevenes, M. A., Pereira, H. & Bermudez, R. Citizen science as a co-creative measure to water quality: chemical data and local participation in a rural territory. *Front. Environ. Sci.* **10**, 940778 (2022).
125. Nardi, F. et al. Citizens and Hydrology (CANDHY): conceptualizing a transdisciplinary framework for citizen science addressing hydrological challenges. *Hydrol. Sci. J.* **67**, 2534–2551 (2022).
126. Dyer, F. et al. Waterwatch data quality: an opportunity to augment professionally collected data sets. In *Proc. 7th Australian Stream Management Conference* 27–30 (ASM, 2014).
127. Rose, L. A., Karwan, D. L. & Godsey, S. E. Concentration–discharge relationships describe solute and sediment mobilization, reaction, and transport at event and longer timescales. *Hydrol. Processes* **32**, 2829–2844 (2018).
128. Hare, D. K., Helton, A. M., Johnson, Z. C., Lane, J. W. & Briggs, M. A. Continental-scale analysis of shallow and deep groundwater contributions to streams. *Nat. Commun.* https://doi.org/10.1038/s41467-021-21651-0 (2021).
129. Li, L. et al. Toward catchment hydro-biogeochemical theories. *WIREs Water* https://doi.org/10.1002/wat2.1495 (2021).
130. Zhi, W. & Li, L. The shallow and deep hypothesis: subsurface vertical chemical contrasts shape nitrate export patterns from different land uses. *Environ. Sci. Technol.* **54**, 11915–11928 (2020).
131. Zhi, W., Klingler, C., Liu, J. & Li, L. Widespread deoxygenation in warming rivers. *Nat. Clim. Change* **13**, 1105–1113 (2023).
132. Li, L. et al. Climate controls on river chemistry. *Earths Future* **10**, e2021EF002603 (2022).
133. Harari, Y. N. *Sapiens: A Brief History of Humankind* (Random House, 2014).
134. Popper, K. *The Logic of Scientific Discovery* (Basic Books, 1959).
135. Hodges, A. *Alan Turing: The Enigma: The Centenary Edition* (Princeton Univ. Press, 2012).
136. Do, H. X., Gudmundsson, L., Leonard, M. & Westra, S. The Global Streamflow Indices and Metadata Archive (GSIM)—Part 1: the production of a daily streamflow archive and metadata. *Earth Syst. Sci. Data* **10**, 765–785 (2018).
137. Virro, H., Amatulli, G., Kmoch, A., Shen, L. & Uuemaa, E. GRQA: global river water quality archive. *Earth System Sci. Data* **13**, 5483–5507 (2021).
138. Gunn, M. A., Matherne, A. M. & Mason, J. R. R. *The USGS at Embudo, New Mexico: 125 Years of Systematic Streamgaging in the United States* Report No. 2014-30344 (USGS, 2014).
139. Burt, T. P. & McDonnell, J. J. Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resour. Res.* **51**, 5919–5928 (2015).
140. Read, E. K. et al. Water quality data for national-scale aquatic research: the Water Quality Portal. *Water Resour. Res.* **53**, 1735–1745 (2017).

141. Council, N. R. *Confronting the Nation's Water Problems: The Role of Research* (National Academies Press, 2004).
142. Li, Z., Liu, H., Zhang, C. & Fu, G. Generative adversarial networks for detecting contamination events in water distribution systems using multi-parameter, multi-site water quality monitoring. *Environ. Sci. Ecotechnol.* **14**, 100231 (2023).
143. Qu, H. & Yuan, W. Water quality Anomaly detection based on optimally reconfigured convolutional autoencoder. In *2022 International Conference on Wearables, Sports and Lifestyle Management (WSLM)* 137–141 (IEEE, 2022).
144. Shen, C., Chen, X. & Laloy, E. Broadening the use of machine learning in hydrology. *Front. Water* https://doi.org/10.3389/frwa.2021.681023 (2021).
145. Schmidhuber, J. Annotated history of modern AI and deep learning. Preprint at https://arxiv.org/abs/2212.11279 (2022).
146. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
147. Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386 (1958).
148. Amari, S.-I. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Trans. Comput.* **100**, 1197–1206 (1972).
149. Maier, H. R. & Dandy, G. C. The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* **32**, 1013–1022 (1996).
150. Maier, H. R. & Dandy, G. C. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Softw.* **15**, 101–124 (2000).
151. Chang, F. J. & Hwang, Y. Y. A self-organization algorithm for real-time flood forecast. *Hydrol. Process.* **13**, 123–138 (1999).
152. Dawson, C. W. & Wilby, R. L. A comparison of artificial neural networks used for river flow forecasting. *Hydrol. Earth Syst. Sci.* **3**, 529–540 (1999).
153. Cigizoglu, H. K. Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. *Adv. Water Res.* **27**, 185–195 (2004).
154. Dransfeld, S., Tatnall, A. R., Robinson, I. S. & Mobley, C. D. A comparison of multi-layer perceptron and multilinear regression algorithms for the inversion of synthetic ocean colour spectra. *Int. J. Remote Sens.* **25**, 4829–4834 (2004).
155. Pankiewicz, G. S. Neural network classification of convective airmasses for a flood forecasting system. *Int. J. Remote Sens.* **18**, 887–898 (1997).
156. Addor, N., Newman, A. J., Mizukami, N. & Clark, M. P. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci.* **21**, 5293–5313 (2017).
157. Newman, A. et al. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrol. Earth Syst. Sci.* **19**, 209 (2015).
158. Kratzert, F. et al. Caravan—a global community dataset for large-sample hydrology. *Sci. Data* https://doi.org/10.1038/s41597-023-01975-w (2023).
159. *GEMStat Database of the Global Environment Monitoring System for Freshwater (GEMS/Water) Programme* (UN Environment Programme, 2018).
160. Hartmann, J., Lauerwald, R. & Moosdorf, N. GLORICH—global river chemistry database. *Pangaea* **902360**, 520 (2019).
161. Rotteveel, L., Heubach, F. & Sterling, S. M. The Surface Water Chemistry (SWatCh) database: a standardized global database of water chemistry to facilitate large-sample hydrological research. *Earth Syst. Sci. Data* **14**, 4667–4680 (2022).
162. Sterle, G. et al. CAMELS-Chem: augmenting CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) with atmospheric and stream water chemistry data. *Hydrol. Earth Syst. Sci. Discuss.* **2022**, 1–23 (2022).
163. D'Alimonte, D. & Zibordi, G. Phytoplankton determination in an optically complex coastal region using a multilayer perceptron neural network. *IEEE Trans. Geosci. Remote Sens.* **41**, 2861–2868 (2003).

Acknowledgements

W.Z. was supported by the National Natural Science Foundation of China (52121006) and by the Barry and Shirley Isett Professorship (to L.L.) at Penn State University. L.L. was supported by the US National Science Foundation via the Critical Zone Collaborative Network (EAR-2012123 and EAR-2012669), Frontier Research in Earth Sciences (EAR-2121621), Signals in Soils (EAR-2034214), and US Department of Energy Environmental System Science (DE-SC0020146). J.P. was supported by Swiss Agency for Development and Cooperation (SDC) (WABES project, 7F-09963.02.01). This paper has been reviewed in accordance with the US Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement or recommendation for use by the US Government. Statements in this publication reflect the authors' professional views and opinions and should not be construed to represent any determination or policy of the US Environmental Protection Agency.

Author contributions

W.Z. and L.L. conceived the idea for the review paper and wrote the first draft. A.P.A., H.E.G. and J.P. provided content for multiple sections and edited multiple versions of the paper. L.L. finalized the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44221-024-00202-z.

Correspondence and requests for materials should be addressed to Li Li.

Peer review information *Nature Water* thanks Danlu Guo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.