A Machine Learner's Guide to Streamflow Prediction

Martin Gauch,[1,2*] **Daniel Klotz,**[2] **Frederik Kratzert,**[2] **Grey Nearing,**[3]
Sepp Hochreiter,[2] **and Jimmy Lin**[1]
[1]David R. Cheriton School of Computer Science, University of Waterloo, ON, Canada
[2]Institute for Machine Learning, Johannes Kepler University Linz, Austria
[3]Google Research, Mountain View, CA, USA

Abstract

Although often subconsciously, many people deal with water-related issues on a daily basis. For instance, many regions rely on hydropower plants to produce their electricity, and, at the extreme, floods and droughts pose one of the big environmental threats of climate change. At the same time, many machine learning researchers have started to look beyond their field and wish to contribute to environmental issues of our time. The modeling of streamflow—the amount of water that flows through a river cross-section at a given time—is a natural starting point to such contributions: It encompasses a variety of tasks that will be familiar to machine learning researchers, but it is also a vital component of flood and drought prediction (among other applications). Moreover, researchers can draw upon large open datasets, sensory networks, and remote sensing data to train their models. As a getting-started resource, this guide provides a brief introduction to streamflow modeling for machine learning researchers and highlights a number of possible research directions where machine learning could advance the domain.

1 Introduction

In the past years, the machine learning community has seen a surge of work around "AI against climate change" [23], "AI for earth sciences", and, more broadly, "AI for social good" [26]. Yet, machine learning researchers starting to work on these topics often face a steep learning curve when it comes to learning the ropes of a domain. In addition, it is often non-trivial to find areas where machine learning can truly contribute to the state of the art. In effect, one has to learn a new "language", as the new domain uses different words for the same thing, the same words for different things, and new words for entirely new concepts. This paper provides a getting-started guide to machine learning researchers who would like to contribute to a field that relates to all the matters we mentioned above: hydrology, the science of water. To stay concise, we restrict the scope of this guide to the prediction of what hydrologists call *streamflow*—the amount of water that flows through a river cross-section at a given time. Accurate streamflow models are, for example, important to manage dams, reservoirs, and hydropower operations, to assess risk of flooding, or to direct help when a flood actually happens.

The contributions of this paper are two-fold: First, we give an introduction to streamflow prediction geared to machine learning researchers. Second, we describe related open research questions that we consider well suited for machine learning contributions.

2 What is Streamflow Prediction?

At its core, streamflow prediction is a spatio-temporal problem. The input data are meteorological variables (e.g., air temperature, precipitation) and geophysical variables (e.g., land cover, soil,

[*]Corresponding author: Martin Gauch, `gauch@ml.jku.at`

AI for Earth Sciences Workshop at NeurIPS 2020.

elevation). During training (but not necessarily during inference), we additionally have observed streamflow target values. The temporal resolution of inputs and predictions vary for different applications. For example, during floods, we want models that predict exact peak times at the finest possible resolution (often in the range of 15 minutes to one hour). For long-term climate change analysis, often daily or even monthly predictions suffice. We note that the following two sections are adapted from the "Data Scientist's Guide to Streamflow Prediction", where interested readers can find a more in-depth discussion [8].

2.1 Understanding *What* We're Predicting

Usually, hydrologists are interested in streamflow at geographically prominent points such as a city, an estuary, or the inflow to a lake. These points are referred to as the *outlets*. Given an outlet point, we can delineate the *upstream area* from which all water drains towards the outlet. This upstream area has several synonymous names: *basin*, *catchment*, or *watershed*. We call a basin *gauged* if its outlet is equipped with a measurement station, which produces a time series of actual streamflow (known as *hydrograph*). Conversely, basins without gauging station are called *ungauged*.

The delineation of basins is largely based on topological information, as (with some exceptions) this defines the downward gradients that the water follows. The size of a basin depends on the selected outlet: the further upstream along a river, the smaller the basin. As rivers and lakes drain into each other, basins contain nested *sub-basins*. The example in Figure 1 illustrates these concepts: The solid black line outlines an ungauged basin, whereas the gauge station at the city represents the outlet of a gauged sub-basin (dashed line). More formally, we can think of the river network as a directed graph where each node is a sub-basin (or any other spatial extent, such as a grid cell) and each edge represents a downstream-relationship. For instance, an edge from node A to B indicates that all water flowing out of cell A drains into cell B. The arrows in Figure 1 show a few examples of these relations between grid cells—note how the arrows point to the adjacent cell with the lowest elevation.

Hydrologists distinguish two concepts: *runoff* and *streamflow*. Runoff refers to an amount of water that flows inside or between basins and their building blocks such as hill slopes. Commonly, runoff is expressed in units of length (e.g., millimetres) during a fixed time period such as one day. An intuitive explanation for this notion is the hypothetical depth of water if all outflowing water was equally distributed over the geographical area of interest (e.g., a grid cell, hill slope, or basin). Streamflow on the other hand is the amount of water that flows through the cross-section at a point along a stream. We can think of streamflow as the aggregation of runoff from across the whole basin that drains into a river. Note that runoff is defined for arbitrary areas, even if there is no stream in the area. Hence, we can talk about the runoff in or from an area; this is the amount of water that the area contributes to the streamflow. Streamflow, meanwhile, is only defined at points on a stream or river, so we cannot

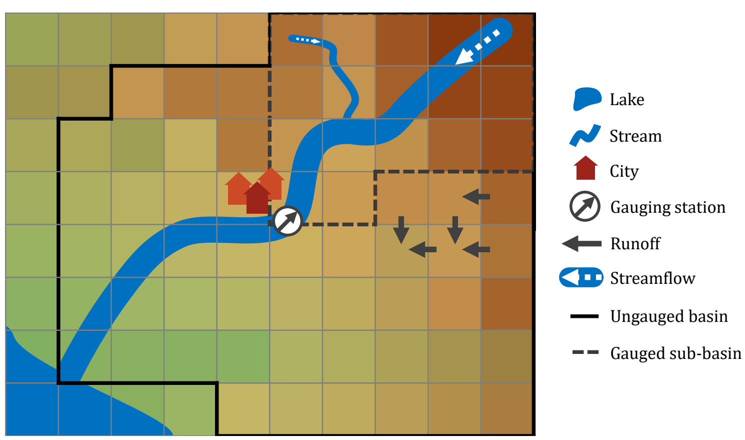

Figure 1: Illustration of streamflow, runoff, basins, and sub-basins. The background gradient represents elevation (brown = high, green = low). From the elevation, we can derive the outline for the ungauged basin (solid black line). The highlighted sub-basin is gauged near a city (dashed line). Unlike streamflow, runoff has a meaning at every grid cell, and the cells are connected through downstream-relationships (gray arrows as a few examples).

talk about the streamflow at arbitrary points in space. Many hydrologic models actually calculate runoff, which is then converted to streamflow by a *routing model* that specifies how fast the runoff accumulates to streamflow along the edges of a river network. Thus, models are often referred to as *rainfall–runoff models*, even though they are usually evaluated on streamflow.

2.2 Understanding *How* We're Predicting

There are two commonly used types of input data: forcings and geophysical data. Forcings are time series of meteorological data such as precipitation and temperature. Their name stems from the fact that these data are needed to run, or *force*, the model. Geophysical data provide information such as land cover, soil, or elevation. Most models assume that these variables are static (read: non-time-series). They are not *actually* static, but they change so slowly that it is often acceptable to treat them as fixed except in long-term simulations. Although these two categories are the two most commonly used, there exist a variety of other possible data sources, such as dam operation schedules or annually recurrent information (e.g., leaf canopy, sunrise, and sunset).

Traditionally, there are three types of models: *conceptual* and *process-based models*, which, to varying degree, both try to simulate the physical processes that generate runoff from rainfall (such as evaporation, transpiration, or soil infiltration), and *data-driven models*, which do not explicitly model physical relationships. Appendix C describes this distinction in more detail.

So far, most machine learning models for streamflow prediction operate in a *lumped* setting: their input forcings and geophysical data are aggregated to a basin-level. A lumped model that predicts streamflow for a certain basin will need "pointwise" forcing time series and geophysical data. For example, a lumped model of the gauged sub-basin in Figure 1 might take as input the sub-basin's minimum and maximum temperature, the cumulative precipitation, and the mean elevation calculated across all cells within the dashed sub-basin outline.

Most traditional hydrologic models are trained and evaluated for a single basin or a fixed set of basins. Although these models try to implement some of the underlying physical processes, in most cases they do not generalize to basins that were not part of the training procedure, a process known as *spatial validation* [14]. Instead, these models are only evaluated in *temporal validation*, where we use the same basins as in training, but a different time period. Temporal validation is much easier than spatial validation, as the relationships between forcings and runoff remain relatively static over time, while the geophysical properties of different basins can result in completely different relationships. More sophisticated models try to adapt their streamflow generation to the geophysical properties of a basin to achieve some degree of spatial generalization. Such spatial generalization is important towards models that can predict ungauged basins (for which no training data exists), which is one of the big open challenges in hydrology [5]. In the context of temporal validation, note that there is some distinction between models that predict future vs. models that predict past streamflow; Appendix B provides more details on this.

Given time series of streamflow ground truth and predictions, there are a variety of metrics to quantify the error [11]. A very common metric (albeit not without criticism [10]) is the *Nash–Sutcliffe efficiency* (NSE), which is defined as the coefficient of determination R^2 of the observed and predicted streamflow [19]. Besides metrics, hydrologists also assess model performance by comparing the *hydrologic signatures* of prediction and observation [17]. These signatures are statistics calculated from the time series of streamflow and provide a more interpretable understanding of model performance.

3 How can Machine Learning Contribute to Hydrology?

In the past few years, deep learning techniques have made their way into research and practical applications of streamflow prediction. Kratzert et al. demonstrated how LSTMs trained on hundreds of basins outperform traditional conceptual models [16] and provide state-of-the-art predictions in ungauged basins (i.e., applied to basins they were not trained on) [15].

Despite these advances, there remain a number of open challenges that may be tackled with machine learning techniques. For hydrologically well-versed readers, we additionally refer to Nearing et al., who provide an (arguably opinionated) analysis of the "State of the Union" of hydrology in the context of recent advances in deep learning [20].

Spatially-Distributed Modeling. As mentioned above, most current machine learning-based rainfall–runoff models operate on lumped basins, i.e., on data that are aggregated across the whole basin. In large basins that span hundreds of thousands square kilometers, however, it matters *where* within the basin the rain falls. One challenge in spatially-distributed modeling is the fact that, unlike in image processing tasks, the target data is very sparsely available in space—we only know the actual streamflow at gauging stations. Another challenge is the irregular shape of basins: unlike images, basins are no rectangles but arbitrary polygons in space.

One potential solution to these problems may be the use of graph networks. As large basins consist of smaller sub-basins (which, in turn, consist of even smaller units) that are connected in a tree structure, it seems natural to model large river networks as small lumped nodes that are connected through downstream-relations. Initial studies in this application show promising results [12, 18], but there remains much to be explored, such as the question whether these models generalize to previously unseen river networks.

Non-Stationarity. Climate change will alter the weather patterns, which in turn will alter the streamflow response [22]. Current models, however, have only seen past data where the climate was relatively stable, compared to what is expected in the future [22]. In effect, the models will have to extrapolate beyond each individual basin's known characteristics. The ability of machine learning models to leverage large datasets that can span continents or even the whole world raises hopes: often, the new climate of a region A will be similar to the current climate in another region B. Recent results with models on large numbers of basins suggest that machine learning models may be able to transfer their trained knowledge from region B to the out-of-sample situation in region A (hydrologists sometimes refer to such strategies as *trading space for time*) [9].

Counterfactual Analysis. The non-stationarity of climate conditions, but also human intervention in ecosystems around the world make it increasingly important for water managers to conduct counterfactual analyses. These analyses could answer questions such as, "*What happens to a basin's streamflow if a forest is cut down?*", "*What if a hydropower dam is built along a river?*", or, "*What if a lake in the basin dries up?*" One way to answer these counterfactual questions might be a simulation environment where researchers can model their region of interest in a controlled fashion, from river network across geophysical landscape properties to climate conditions.

Reinforcement Learning and Multi-Agent Learning. The words "simulation environment" should already have drawn the attention of reinforcement learning researchers, and indeed, we do see applications for this area of machine learning. One example is the operation of hydropower plants and dams: the decision when to open or close a dam is embedded in a complex network of multi-party decision making. Operators need to consider the capacity of the reservoir together with the current and future demand, supply, and price of the produced energy, and they need to make sure the downstream river sections neither dry up while the dam is closed nor overflow while it is opened.

Inductive Bias. This branch of research tries to incorporate known constraints and information about the target domain into machine learning models. For models of physical systems, this paradigm is also known as physics- or theory-guided machine learning [13]. One manifestation of inductive biases in rainfall–runoff modeling is the idea of designing a model in a graph that stems from the spatial structure of the river network. Another example are Mass-Conserving LSTMs (MC-LSTMs), a recently proposed variant of LSTMs, where *mass inputs* are conserved by design [3]. The authors proved the suitability of MC-LSTM for rainfall–runoff modeling, and, while slightly worse than LSTM in terms of NSE performance, the model achieves new state-of-the-art for peak flow prediction. At the same time, however, it is often surprisingly hard to frame prior knowledge in a way that benefits the machine learning model. For instance, Frame et al. showed that the output of a process-based rainfall–runoff model does not significantly improve the predictions of an LSTM [7].

Finally, we would like to close by emphasizing that we consider it extremely important to conduct interdisciplinary research in close collaboration with domain experts who can assess the contributions from the viewpoint of the application. Nevertheless, we hope that this introductory guide acts as a helpful resource for machine learning researchers who start to work on related topics in hydrology and, since many of the concepts apply more broadly, environmental sciences.

Acknowledgments and Disclosure of Funding

We thank Shervan Gharari from the University of Saskatchewan and Juliane Mai from the University of Waterloo for their valuable comments and suggestions throughout the creation of this paper. Further, we thank the Global Water Futures program and the Canada First Research Excellence Fund for their financial support.

References

[1] N. Addor, A. Newman, N. Mizukami, and M. P. Clark. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10):5293–5313, 2017.

[2] C. Alvarez-Garreton, P. A. Mendoza, J. P. Boisier, N. Addor, M. Galleguillos, M. Zambrano-Bigiarini, A. Lara, C. Puelma, G. Cortes, R. Garreaud, J. McPhee, and A. Ayala. The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset. *Hydrology and Earth System Sciences*, 22(11):5817–5846, 2018.

[3] Anonymous. MC-LSTM: Mass-conserving LSTM. In *Submitted to International Conference on Learning Representations*, 2021. under review.

[4] K. Beven and P. Young. A guide to good practice in modeling semantics for authors and referees. *Water Resources Research*, 49(8):5092–5098, 2013.

[5] G. Blöschl, M. F. P. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, J. W. Kirchner, J. J. McDonnell, H. H. Savenije, M. Sivapalan, C. Stumpp, E. Toth, E. Volpi, G. Carr, C. Lupton, J. Salinas, B. Széles, A. Viglione, H. Aksoy, S. T. Allen, A. Amin, V. Andréassian, B. Arheimer, S. K. Aryal, V. Baker, E. Bardsley, M. H. Barendrecht, A. Bartosova, O. Batelaan, W. R. Berghuijs, K. Beven, T. Blume, T. Bogaard, P. B. de Amorim, M. E. Böttcher, G. Boulet, K. Breinl, M. Brilly, L. Brocca, W. Buytaert, A. Castellarin, A. Castelletti, X. Chen, Y. Chen, Y. Chen, P. Chifflard, P. Claps, M. P. Clark, A. L. Collins, B. Croke, A. Dathe, P. C. David, F. P. J. de Barros, G. de Rooij, G. D. Baldassarre, J. M. Driscoll, D. Duethmann, R. Dwivedi, E. Eris, W. H. Farmer, J. Feiccabrino, G. Ferguson, E. Ferrari, S. Ferraris, B. Fersch, D. Finger, L. Foglia, K. Fowler, B. Gartsman, S. Gascoin, E. Gaume, A. Gelfan, J. Geris, S. Gharari, T. Gleeson, M. Glendell, A. G. Bevacqua, M. P. González-Dugo, S. Grimaldi, A. B. Gupta, B. Guse, D. Han, D. Hannah, A. Harpold, S. Haun, K. Heal, K. Helfricht, M. Herrnegger, M. Hipsey, H. Hlaváčiková, C. Hohmann, L. Holko, C. Hopkinson, M. Hrachowitz, T. H. Illangasekare, A. Inam, C. Innocente, E. Istanbulluoglu, B. Jarihani, Z. Kalantari, A. Kalvans, S. Khanal, S. Khatami, J. Kiesel, M. Kirkby, W. Knoben, K. Kochanek, S. Kohnová, A. Kolechkina, S. Krause, D. Kreamer, H. Kreibich, H. Kunstmann, H. Lange, M. L. R. Liberato, E. Lindquist, T. Link, J. Liu, D. P. Loucks, C. Luce, G. Mahé, O. Makarieva, J. Malard, S. Mashtayeva, S. Maskey, J. Mas-Pla, M. Mavrova-Guirguinova, M. Mazzoleni, S. Mernild, B. D. Misstear, A. Montanari, H. Müller-Thomy, A. Nabizadeh, F. Nardi, C. Neale, N. Nesterova, B. Nurtaev, V. O. Odongo, S. Panda, S. Pande, Z. Pang, G. Papacharalampous, C. Perrin, L. Pfister, R. Pimentel, M. J. Polo, D. Post, C. P. Sierra, M.-H. Ramos, M. Renner, J. E. Reynolds, E. Ridolfi, R. Rigon, M. Riva, D. E. Robertson, R. Rosso, T. Roy, J. H. Sá, G. Salvadori, M. Sandells, B. Schaefli, A. Schumann, A. Scolobig, J. Seibert, E. Servat, M. Shafiei, A. Sharma, M. Sidibe, R. C. Sidle, T. Skaugen, H. Smith, S. M. Spiessl, L. Stein, I. Steinsland, U. Strasser, B. Su, J. Szolgay, D. Tarboton, F. Tauro, G. Thirel, F. Tian, R. Tong, K. Tussupova, H. Tyralis, R. Uijlenhoet, R. van Beek, R. J. van der Ent, M. van der Ploeg, A. F. V. Loon, I. van Meerveld, R. van Nooijen, P. R. van Oel, J.-P. Vidal, J. von Freyberg, S. Vorogushyn, P. Wachniew, A. J. Wade, P. Ward, I. K. Westerberg, C. White, E. F. Wood, R. Woods, Z. Xu, K. K. Yilmaz, and Y. Zhang. Twenty-three unsolved problems in hydrology (UPH) — a community perspective. *Hydrological Sciences Journal*, 64(10):1141–1158, 2019.

[6] G. Coxon, N. Addor, J. P. Bloomfield, J. Freer, M. Fry, J. Hannaford, N. J. K. Howden, R. Lane, M. Lewis, E. L. Robinson, T. Wagener, and R. Woods. Catchment attributes and hydro-meteorological timeseries for 671 catchments across Great Britain (CAMELS-GB), 2020.

[7] J. Frame, G. Nearing, F. Kratzert, and M. Rahman. Post processing the U.S. National Water Model with a long short-term memory network, 2020.

[8] M. Gauch and J. Lin. A Data Scientist's guide to streamflow prediction. 2020.

[9] M. Gauch, J. Mai, and J. Lin. The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. *Environmental Modelling & Software*, 135:104926, 2021.

[10] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91, 2009.

[11] H. V. Gupta, S. Sorooshian, and P. O. Yapo. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4):751–763, 1998.

[12] X. Jia, J. Zwart, J. Sadler, A. Appling, S. Oliver, S. Markstrom, J. Willard, S. Xu, M. Steinbach, J. Read, and V. Kumar. Physics-guided recurrent graph networks for predicting flow and temperature in river networks. 2020.

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

[14] V. Klemeš. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1):13–24, 1986.

[15] F. Kratzert, D. Klotz, M. Herrnegger, A. K. Sampson, S. Hochreiter, and G. Nearing. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12):11344–11354, 2019.

[16] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing. 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, 2019.

[17] H. McMillan. Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, 34(6):1393–1409, 2020.

[18] Z. Moshe, A. Metzger, G. Elidan, F. Kratzert, S. Nevo, and R. El-Yaniv. HydroNets: Leveraging river structure for hydrologic modeling. 2020.

[19] J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part I—a discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970.

[20] G. Nearing, F. Kratzert, A. K. Sampson, C. Pelissier, D. Klotz, J. Frame, and H. Gupta. What role does hydrological science play in the age of machine learning? 2020.

[21] A. Newman, K. Sampson, M. P. Clark, A. Bock, R. Viger, and D. Blodgett. A large-sample watershed-scale hydrometeorological dataset for the contiguous USA, 2014.

[22] R. K. Pachauri and L. Meyer, editors. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 2014.

[23] D. Rolnick, P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, A. S. Ross, N. Milojevic-Dupont, N. Jaques, A. Waldman-Brown, A. Luccioni, T. Maharaj, E. D. Sherwin, S. K. Mukkavilli, K. P. Kording, C. Gomes, A. Y. Ng, D. Hassabis, J. C. Platt, F. Creutzig, J. Chayes, and Y. Bengio. Tackling climate change with machine learning. 2019.

[24] J. Sitterson, C. Knightes, R. Parmar, K. Wolfe, M. Muche, and B. Avant. An overview of rainfall-runoff model types. *U.S. Environmental Protection Agency*, EPA/600/R-17/482, 2017.

[25] E. Todini. Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences Discussions*, 11(1):468–482, 2007.

[26] N. Tomašev, J. Cornebise, F. Hutter, S. Mohamed, A. Picciariello, B. Connelly, D. C. Belgrave, D. Ezer, F. C. van der Haert, F. Mugisha, et al. AI for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1):1–6, 2020.

A Practical Notes on Getting Started with Streamflow Prediction

For researchers who would like to explore streamflow prediction more practically, we have implemented the `neuralhydrology` Python package.[2] This library provides code to load data, train, and evaluate pre-packaged or custom-built machine learning models for various streamflow prediction tasks. One common corpus of datasets are the *Catchment Attributes and Meteorology for Large-sample Studies* (CAMELS) that contain meteorological forcings and geophysical characteristics for basins across the continental United States [1, 21]. More recently, similar datasets for basins in Chile, CAMELS-CL [2], and Great Britain, CAMELS-GB [6], have been published, and datasets for other regions are in development. It is important to note, however, that each of these datasets is based on its own meteorological forcing product. Consequently, the resulting lumped datasets have different biases and we cannot simply apply a model that was, for example, trained on CAMELS-US to basins from the Chilean version.

B Forecasting vs. Simulation

If we predict the future, we have to rely on forcings that are predictions themselves (or not use future forcings at all). These future data can, for instance, come from a numerical weather prediction model. For simulations of historical streamflow, the forcings can, if available, be observed data. Since observations are generally sparse and subject to large uncertainties, however, we commonly use forcings that are generated by other models (e.g., by climate reanalyses, a type of data product that combines different model simulations and measurements to provide a good estimate of the historical meteorologic variables). In this context, note the distinction between *simulation* models that predict *past* streamflow and *forecasting* models that predict streamflow for *future* dates. Of course, to develop a forecasting model, we can always act as if the input were only available up to a certain time step in the past, and then "hindcast" the past streamflow after this time step. We refer to Beven and Young [4] for a more fine-grained distinction between different prediction modes.

C Types of Hydrologic Models

Traditionally, there are three broad classes of models used by hydrologists [24]:

Conceptual models are simplified representations based on pseudo-physical understanding of catchment processes. Often, these are "leaky bucket"-type models where storage tanks model water stores, sometimes with small nonlinearities included in key places like infiltration and evapotranspiration. The parameters of these models are sometimes conceptually linked to geophysical properties of a watershed, but accurate predictions require parameter estimation. Thus, conceptual models must (albeit with many exceptions) be calibrated [25]. This is done using standard optimization techniques, and in many ways is like training the weights and biases of a deep learning model.

Process-based models attempt to explicitly represent catchment processes using bio-geophysical understanding, often motivated by the idea that predicting physical systems under changing conditions requires understanding causal mechanisms. Thus, the parameters of process-based models are generally intended to be explicit geophysical properties, and the goal is often to do as little calibration as possible by accurately mapping catchment properties.

Data-driven models lack almost any physical understanding, and there has historically been resistance in the hydrology community to adopting this type of model in practice [5].

As a general rule of thumb, data-driven models are more accurate than conceptual models, which are more accurate than process-based models [25].

[2] `https://github.com/neuralhydrology/neuralhydrology`