Environmental Data Science (2025), 4: e5, 1–15
doi:10.1017/eds.2024.48

CAMBRIDGE
UNIVERSITY PRESS

APPLICATION PAPER

Streamflow prediction using artificial neural networks and soil moisture proxies

Robert Edwin Rouse[1], Doran Khamis[2], Scott Hosking[3,4], Allan McRobie[1] and Emily Shuckburgh[5]

[1]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
[2]UK Centre for Ecology & Hydrology, Wallingford OX10 8BB, UK
[3]British Antarctic Survey, Cambridge CB3 0ET, UK
[4]The Alan Turing Institute, London NW1 2DB, UK
[5]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, UK
Corresponding author: Robert Edwin Rouse; Email: rer44@cam.ac.uk

Received: 30 October 2023; **Revised:** 04 September 2024; **Accepted:** 16 September 2024

Keywords: artificial neural networks; hydrology; machine learning; streamflow

Abstract

Machine learning models have been used extensively in hydrology, but issues persist with regard to their transparency, and there is currently no identifiable best practice for forcing variables in streamflow or flood modeling. In this paper, using data from the Centre for Ecology & Hydrology's National River Flow Archive and from the European Centre for Medium-Range Weather Forecasts, we present a study that focuses on the input variable set for a neural network streamflow model to demonstrate how certain variables can be internalized, leading to a compressed feature set. By highlighting this capability to learn effectively using proxy variables, we demonstrate a more transferable framework that minimizes sensing requirements and that enables a route toward generalizing models.

Impact Statement

This paper addresses the challenge of developing universally transferable hydrological modeling approaches using machine learning by investigating the feature space and demonstrating the case for more ubiquitous physics-based proxy variables. The authors propose an approach that utilizes only meteorological variables to force a streamflow model with proxies for measurements of internal mechanics without a loss of predictive performance. Such an approach could enable more effective general hydrological machine learning models to assess climate change impact on global hydrological systems.

1. Introduction

Artificial neural aetworks (ANNs) in various forms have been studied in hydrology over a significant period of time with a high level of model skill in many cases (Aichouri et al., 2015; Ali and Shahbaz, 2020; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Ayzel and Heistermann, 2021; Dawson et al., 2006; Gao et al., 2020; Govindaraju et al., 2000; Kumar et al., 2004). However, they have not been widely taken up for operational use, perhaps due to the perceived lack of interpretability of

This research article was awarded Open Materials badge for transparent practices. See the Data Availability Statement for details.

model mechanics, and a lack of consistency in data frameworks and model specification (Abrahart et al., 2012). When compared with established mechanistic models, such as the Système Hydrologique Européen (SHE) (Abbott et al., 1986a, 1986b) or the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998; Srinivasan et al., 1998), which have been used extensively (Refsgaard et al., 2010), there is no widely accepted machine learning framework, despite the relative maturity of the algorithms themselves.

Mechanistic methods requiring extensive calibration and large amounts of hydromorphological data (Devi et al., 2015; Jaiswal et al., 2020) can, therefore, be difficult to apply globally. This is partly due to the inequity of meteorological and hydrological data collection (Kidd et al., 2017; Krabbenhoft et al., 2022) (many flood events have little or no observational data associated with them (Robson and Reed, 2008)), but looking to the future, climate nonstationarity driven by anthropogenic climate change (Wilby and Quinn, 2013) could cause issues in precalibrated models. However, we might perhaps learn patterns from areas of extensive measurement and extrapolate toward those without. It is in this case that machine learning might be best applied (Bishop, 2016).

On other empirical methods, such as those described in the Flood Estimation Handbook (Calver et al., 2009; Faulkner et al., 2012; Institute of Hydrology, 2008; Samuels et al., 2008), our concern is that their utility may diminish in the face of the aforementioned nonstationarity, which may cause a shift in flood event distributions through an increase in the frequency of extreme events, and the tighter demands that might be expected of predictions in response to climate change.

If data availability is a problem for many locations, we might wish to rule out approaches that use measurements of internal processes within a catchment. This could rule out the application of autoregressive approaches that require prior knowledge of runoff or flow values (Aichouri et al., 2015; Ali and Shahbaz, 2020). In this work, we posit that “external” measurements (those that do not require direct observation inside a catchment, for example, gridded output from a meteorological model) are the only ones that we can assume reliably available, making our adopted position an approach that has to be an input–output model but still explainable.

To achieve this, we aim to find a way to represent the internal state of a catchment, in terms of water storage as soil moisture, using more easily attainable “external” variables. The representation of internal soil moisture state we will approximate through statistical parameterizations of the histories of meteorological variables, resulting in what we term antecedent proxies. We use these antecedent proxies to force the machine learning model alongside the other input variables and compare to outputs from a model of the same architecture that takes soil moisture as input instead of the proxies. We hope that this highlights how machine learning models can be used to reframe a problem while remaining faithful to the established understanding of science in the field.

Our goal, therefore, is to lay a foundation for maximizing the generalization capability (in terms of region of applicability) of hydrological machine learning models by minimizing the data burden. We have three objectives. First, we aim to create a model that responds only to “external” climatic variables (those that can be accessed without measurement within the catchment). Second, we aim for model parsimony by balancing the number of input variables against model skill. We base this on the premise that using fewer variables leads to a more explainable model and one that is easier to apply. Third, our variable choices must be physically reasoned. This is particularly true of the proxy variables that we choose to represent those internal processes of a catchment that we do not observe.

2. Theory and methods

2.1. Catchment conceptualization

In terms of the inputs and outputs that will form the basis of a machine learning model, we consider a crude representation of an inland catchment, one we assume is not affected by tidal inflows, with water entering the catchment through precipitation and exiting through evaporation, transpiration, and through the streamflow outflow, our target variable. We think of the internal processes as ones that move water toward one of the outflow processes or into internal storage.

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

More formally, the sum total of water that leaves a catchment via river discharge, that enters or leaves through climatic, or other, processes, and any changes to water stored within a catchment must be equal to zero, as per the integral form of continuity in Equation (2.1)) (Yilmaz et al., 2008), where the cumulative precipitation over a time period with length T in the catchment is given by $\int_{t=0}^{T}\Psi_t dt$, the water leaving the catchment through evapotranspiration is $\int_{t=0}^{T}Edt$, the target streamflow exiting through the river is $\therefore\int_{t=0}^{T}Y_t$, and the change in water storage is ΔS.

$$\begin{aligned} 0 &= \int_{t=0}^{T}\Psi_t dt - \int_{t=0}^{T}Y_t dt - \int_{t=0}^{T}Edt - \Delta S \\ \therefore \int_{t=0}^{T}Y_t dt &= \int_{t=0}^{T}\Psi_t dt - \int_{t=0}^{T}Edt - \Delta S \end{aligned} \tag{2.1}$$

We further simplify the hydrological system by considering the meteorological inputs and outputs as acting at a single point, rather than being spatially distributed. This single point we take is the center of the catchment, or, more concretely, the centroid of the catchment, being the geometric center of a catchment's two-dimensional surface, equivalent to its center of mass for a congruent shape of uniform weight density. Assuming that a catchment is a shape in two dimensions with area, A, its centroid, with coordinates $(\bar{x},\bar{y})$, is given by taking first moments of area in both axes, with respect to some frame of reference, divided by the shape's area, as in Equation (2.2)):

$$\bar{x} = \frac{\int_A xdA}{\int_A dA}; \quad \bar{y} = \frac{\int_A ydA}{\int_A dA} \tag{2.2}$$

where, for the purposes of this calculation, we will consider the catchment as a surface that exists in a flat, two-dimensional plane, effectively removing elevation and reducing the dimensionality of its boundary, B, such that $B \in \mathbb{R}^2$ rather than $B \in \mathbb{R}^3$ and treating that boundary as the continuous edge of an irregular shape with uniform density.

As we focus on catchment-specific models, we therefore do not use catchment descriptors as inputs. The predicted flow dynamics are dependent in a nonlinear way on the contemporaneous catchment descriptors, rather than stationary catchment descriptors, and the precipitation response function is not constant; rather, it is learned as a function of the antecedent conditions, captured through soil moisture or its proxies. Furthermore, many catchment descriptors within the United Kingdom are relatively stable; for example, the total amount of agricultural land, representing 72.8% of the UK's total land cover, only varied by 1.3% in a 30-year period from 1990 to 2020 (Department for Environment, Food, and Rural Affairs, 2024; Department for Levelling Up, Housing and Communities, 2023).

2.2. Artificial neural networks

2.2.1. Overview

The multilayer perceptron (MLP) is composed of multiple, interconnected neuron units, each combining inputs linearly before an activation function is applied; depending on the choice for that activation function, nonlinearity can be introduced. A single neuron along the structure for an arbitrary MLP is shown in Figure 1, where the output, y, from a single neuron with respect to its inputs, $\mathbf{x}$, is a general linear combination of those inputs, with weights $\mathbf{w}$ and bias b, prior to the activation function, α, being applied.

If we let Φ represent the set of all parameters to be learned, then we can use the difference between a prediction made with the network, y', and the target value, y, through a cost function, $J(\Phi)$, such as the root-mean-squared error (RMSE) form in Equation 2.3.

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

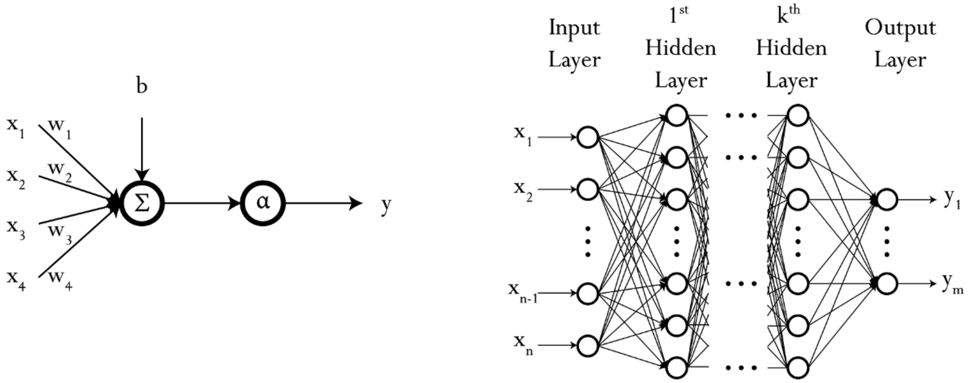


Figure 1. *Graphical representation of a single neuron unit and an arbitrary neural network with l layers.*

$$J(\Phi) = \left(\frac{1}{2n} \sum_{i=1}^{n} \left(y_i' - y_i \right)^2 \right)^{\frac{1}{2}} \tag{2.3}$$

To minimize the value of the cost function, with respect to the parameters Φ, we typically use gradient descent to update each parameter, ϕ_j, according to the sign and magnitude of the gradient, controlled by the learning rate hyperparameter η, as per Equation 2.4.

$$\phi_j := \phi_j - \eta \cdot \frac{\partial J(\Phi)}{\partial \phi_j} \tag{2.4}$$

The Backpropagation algorithm (Rumelhart et al., 1986) enables the application of gradient descent in networks by propagating the derivatives backward between layers; for the parameters at the l^{th} layer of a network with L layers, the update gradient for that layer is chained back through all previous layers and activation functions, $\alpha_L, \dots, \alpha_l$, as in Equation 2.5.

$$\frac{\partial J(\Phi)}{\partial \Phi_l} := \frac{\partial J(\Phi)}{\partial \alpha_L} \cdot \frac{\partial(\alpha_L)}{\partial \Phi_{L-1}} \cdot \dots \cdot \frac{\partial(\alpha_l)}{\partial \Phi_l} \tag{2.5}$$

2.2.2. Implementation

The model we define here is applied individually to each catchment: instances of the model learn from and predict for each catchment separately. The MLP structure is relatively simple, comprising only two hidden layers with unitary output, where the number of nodes in the first hidden layer is half the dimension of the input space and the number of nodes in the second hidden layer is halved again. The activation function is the sigmoid linear unit (Ramachandran et al., 2017), and the optimization algorithm we use is the Adam algorithm, its name coming from adaptive moment estimation (Kingma and Ba, 2017; Ruder, 2017). Inputs to the model are also normalized, using values from the training set only, to smooth training.

An advantage of MLPs is that they offer a simple baseline neural network that, as a universal function approximator, is still capable of learning complex nonlinear relationships. Given that MLPs have demonstrated high predictive skill in extensive application to hydrological problems (Aichouri et al., 2015; Ali and Shahbaz, 2020; ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Ayzel and Heistermann, 2021; Dawson et al., 2006; Gao et al., 2020; Govindaraju et al., 2000; Kumar et al., 2004), this learning capability includes the relationship between the catchment state and the response to precipitative forcing. Furthermore, extracting the sensitivity of the model to individual variables is a straightforward process that enables the analysis of the relative importance of each variable.

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

2.3. Error metrics

To assess the performance of the equation, we employ the following metrics: RMSE, expressed in Equation 2.6; mean percent relative error (MPRE), expressed in Equation 2.7; relative bias (RB), expressed in Equation 2.8; Nash-Sutcliffe efficiency (NSE), expressed in Equation 2.9; and Kling-Gupta efficiency (KGE), expressed in Equation 2.10. The RMSE and MPRE both give an indication of the absolute magnitude of the absolute error and, accordingly, have a range of $[0,\infty)_{,}$ where 0 indicates a perfect score. RB, on the other hand, identifies whether or not a model is on average under- or overpredicting and has a range of $(-\infty,\infty]$. NSE and KGE, through normalization, enable the model's accuracy to be compared across catchments more easily and are commonly used in hydrology to do so (Gupta et al., 2009; Nash and Sutcliffe, 1970), each with a range of $(-\infty,1]$. While both metrics provide a sense of how the model is performing compared to the mean, where an NSE of 0 and a KGE of -0.41 would indicate performance equivalent to the mean flow benchmark (Wouter et al., 2019), KGE is intended to capture the discrepancy between the variability and timing of the compared datasets. While obtaining a score of 1 1 for the NSE and KGE would be ideal, obtaining scores of NSE ≥ 0.5 (Moriasi et al., 2007) and KGE ≥ 0.5 (Rogelis et al., 2016) are thresholds of suitable model performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y'_i\right)^2}{n}} \tag{2.6}$$

$$MPRE = \frac{100}{n}\cdot\sum_{i=1}^{n}\left|\frac{y_i - y'_i}{y_i}\right| \tag{2.7}$$

$$RB = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{y_i - y'_i}{y_i} \tag{2.8}$$

$$NSE = 1 - \frac{\Sigma_{i=1}^{n}\left(y_i - y'_i\right)^2}{\Sigma_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{2.9}$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\overline{y}'}{\overline{y}} - 1\right)^2 + \left(\frac{\sigma_{y'}}{\sigma_y} - 1\right)^2} \tag{2.10}$$

All of these metrics are expressed in terms of the n pairwise observations, y_i, and model predictions, y'_i, with the KGE also relying on the use of the linear correlation between observations and simulations, r, and the standard deviation of predictions, $\sigma_{y'}$, and observations, σ_y.

2.4. Sensitivity and comparative analysis

To make a comparison of the impact that the soil moisture and antecedent proxy variables have, we utilize an additional pair of methodologies. The first is the correlation between variables, specifically the Pearson Correlation Coefficient, r_{xy}, in Equation 2.11, which we use to determine the information similarity between the soil moisture variables and the proxies.

$$r_{x_1x_2} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}} \tag{2.11}$$

The second methodology is on the sensitivity of a network to the input variables, for which various methodologies exist. Commonly employed methods for doing so include, but are not limited to: input perturbation analysis, weight analysis, and gradient analysis (Cao et al., 2016; Cao and Qiao, 2008; Gevrey et al., 2003; Pizarroso et al., 2022). The method that we employ here is the input perturbation

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

algorithm, where for each input variable, x_i, we add some increment, δ_i, that we can vary and then measure the change in output, where δ_i is a proportion of the input value.

3. Data

The UK Centre for Ecology and Hydrology provides extensive records of hydrometric data from across the UK, with 1602 gauge station records available through the UK National River Flow Archive (NRFA) (UK Centre for Ecology and Hydrology, 2022). For this study, we select representative catchments at difference area scales. Our exemplar catchments are the River Severn at Haw Bridge, selected randomly from the largest 0.5% of catchments with an area of 9895 km^2, the Bedford Ouse at Roxton, from the largest 5% of catchments at 1660 km^2, and the Findhorn at Shenachie, from the largest 50% of catchments at 416 km^2. The elevation, land use, and geology characteristics of these catchments are shown in Figure 2.

The three catchments cover highland, lowland, and mixed elevation catchments that also feature a range of different land use scenarios, with the Findhorn at Shenachie being predominantly mountainous, at 74.32% of land cover, and the Bedford Ouse at Roxton featuring a heavy proportion of arable or horticultural farmland and grassland, 51.33% and 29.66%, respectively. The range of geological representation is less pronounced but still significant, with the Findhorn at Shenachie being 100% very low permeability bedrock and the other two catchments having more mixed compositions.

The input meteorological data are extracted from The European Centre for Medium-Range Weather Forecast's (ECMWF) fifth generation of global climate reanalysis data, ERA5, a modeled data product that utilizes observational data through data assimilation (Hersbach et al., 2018). This data product is widely used, including in hydrological studies, due to its accuracy and comparability with observations (Hersbach et al., 2020; Nogueira, 2020; Tarek et al., 2020). Soil moisture from ERA5, critical to this study, has also been observed to have a high degree of correlation with observations globally, more so than other data products (Dong et al., 2022; Lal et al., 2022; Li et al., 2020). Although the spatial resolution of ERA5 is relatively coarse, at a grid size of 31 km, its ubiquity is highly compatible with our overarching aim of

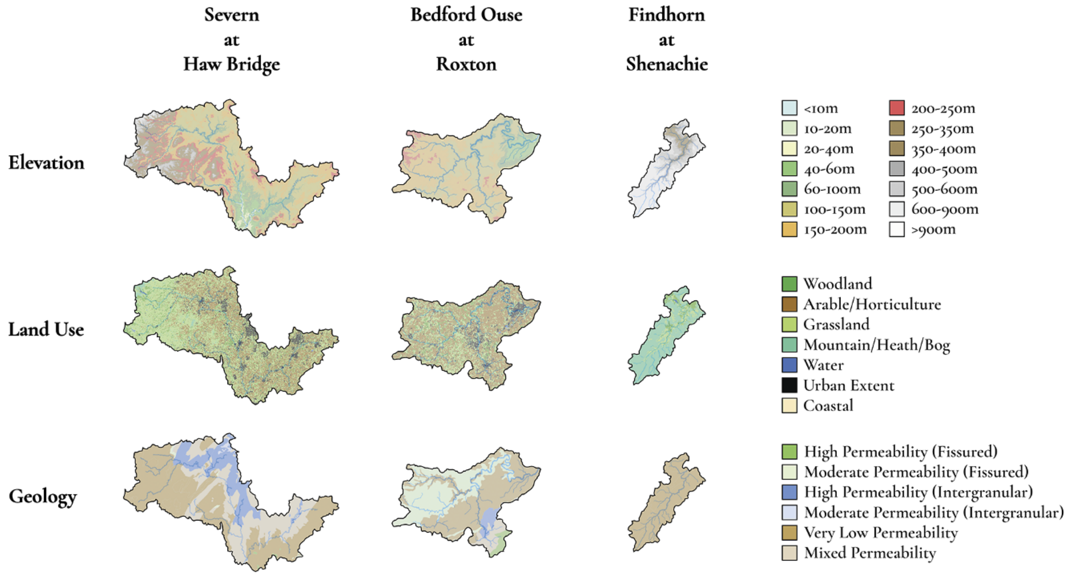


Figure 2. *Elevation, land use, and geology maps for the three study catchments; corresponding keys represent the proportion of each catchment that falls under the respective subcategories. Note that the catchments are shown at different scales for visibility. Adapted from the National River Flow Archive (UK Centre for Ecology and Hydrology, 2022).*

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

enabling a globally applicable, physics-informed machine learning approach, one for transfer to regions where rainfall and soil moisture gauges are not routinely, or at all, available (Kidd et al., 2017; Krabbenhoft et al., 2022).

The input variables used are precipitation at the surface, four levels of soil moisture, and the variables that affect evapotranspiration: daily average temperature, relative humidity, and the components of wind speed parallel to the earth's surface, all of which have been taken at a height just above the surface at 1000 hPa. Up to 28 days worth of meteorological data are used per prediction. We also define a set of proxy variables that serve as a replacement for soil moisture inputs. To encapsulate the antecedent conditions in the catchment and to capture seasonal trends we use the 30, 90, and 180 day moving averages for precipitation and temperature.

Catchment boundaries are provided through the NRFA, derived from the Centre for Ecology & Hydrology's Integrated Hydrological Digital Terrain Model (Morris and Flavin, 1990, 1994), with the centroid calculated as we describe in Section 2. Meteorological data are then interpolated at this point using a cubic spline fitting method (Balog et al., 2023; De Boor, 2001; Tabor and Williams, 2010).

The streamflow data that form our target output are mean daily gauged data. Though these data are readily available, streamflow response to climatic conditions is smoothed at the daily time step, with the maximum instantaneous peak flow not fully represented (Bartens and Haberlandt, 2021; Ding et al., 2015; Fill and Steiner, 2003). This is potentially a greater issue for smaller catchments, where the time for runoff to reach the river is similarly small, causing the instantaneous peak flow to diverge considerably from the mean daily flow (Fill and Steiner, 2003; Fuller, 1914; Gray, 1973). While this issue is apparent, it is not something that we intend to address here and does not diminish the comparison we make in this study.

Finally, the split between training and test data is kept consistent across all three model applications, with the period from 1979 to 2008 used to train model instances, the year 2009 used as a validation set, and the period 2010 to 2019 used to test the model.

4. Results and discussion

4.1. Variable comparison

We first provide analysis of the correlation between the antecedent proxies with the soil moisture variables, in accordance with the methods as described in Section 2, to highlight the information value of these variables and the potential utility of the antecedent proxies. The correlation between soil moisture and the antecedent proxies is shown in Table 1, where SM_{Lx} is the soil moisture at the level x and $\mu_{v:d}$ is the antecedent proxy for a variable v, either precipitation or temperature, taken over a number of days d.

Key to the premise of this work is the strength of correlation between the antecedent proxies and the soil moisture levels. The results are intuitive: shorter-term antecedent proxies have high correlation (positive for precipitation and negative for temperature) with the shallow soil moisture levels; longer-term antecedent proxies exhibit high correlation with the deeper soil moisture levels. It is clear that different information is contained within the antecedent proxies across the timescales. Because we can reason physically that both shallow and deep soil moisture states are important for flow prediction (though the weighting of importance will be catchment dependent) due to their relationship to surface runoff and groundwater recharge, we are justified in selecting antecedent proxy timescales that correlate strongly with both deep and shallow soil moisture states.

4.2. Model results

Our results for the three catchments across all metrics are shown in Table 2, respectively, including results for the models using a meteorological input sequence length of 28 days, 7 days, 7 days with soil moisture, and 7 days with proxy variables. The performance measured by RMSE and MPRE indicates that the highest performing models are those that use the 7-day meteorological input with the inclusion of either the antecedent proxies or soil moisture variables. Increasing the number of days in the meteorological input space continued to increase performance, although we found that the gains in performance started to

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

Table 1. *Pearson correlation coefficients between soil moisture-level variables and antecedent proxies, averaged over the three study catchments*

Variable	SM_{L1}	SM_{L2}	SM_{L3}	SM_{L4}	$\mu_{\mathrm{p:30}}$	$\mu_{\mathrm{p:90}}$	$\mu_{\mathrm{p:180}}$	$\mu_{\mathrm{t:30}}$	$\mu_{\mathrm{t:90}}$	$\mu_{\mathrm{t:180}}$
SM_{L1}	1.00	0.90	0.62	0.16	0.46	0.32	0.27	−0.49	−0.29	0.11
SM_{L2}	0.90	1.00	0.77	0.23	0.50	0.38	0.32	−0.57	−0.37	0.08
SM_{L3}	0.62	0.77	1.00	0.49	0.37	0.39	0.35	−0.68	−0.57	−0.11
SM_{L4}	0.16	0.23	0.49	1.00	0.18	0.40	0.51	−0.45	−0.60	−0.51
$\mu_{p:30}$	0.46	0.50	0.37	0.18	1.00	0.57	0.39	0.01	0.06	0.05
$\mu_{p:90}$	0.32	0.38	0.39	0.40	0.57	1.00	0.71	0.01	0.06	0.10
$\mu_{p:180}$	0.27	0.32	0.35	0.51	0.39	0.71	1.00	−0.08	−0.02	0.06
$\mu_{t:30}$	−0.49	−0.57	−0.68	−0.45	0.01	0.01	−0.08	1.00	0.83	0.29
$\mu_{t:90}$	−0.29	−0.37	−0.57	−0.60	0.06	0.06	−0.02	0.83	1.00	0.72
$\mu_{t:180}$	0.11	0.08	−0.11	−0.51	0.05	0.10	0.06	0.29	0.72	1.00

While the soil moisture-level variables exhibit positive correlation among themselves, the correlation between the deepest level, 4, and the shallowest level, 1, is weak. This suggests that there is significant information gain in utilizing all soil moisture levels simultaneously. (This does not mean that the trained neural network will respond to all variables equally: that quantification is shown in the subsequent sensitivity analysis.)

Table 2. *MLP model performance in terms of Nash-Sutcliffe efficiency for meteorological input sequence length of 28 days (28), 7 days (7), 7 days with soil moisture (7 + SM), and 7 days with proxy variables (7 + P) (bold typeface indicates highest performance)*

		Metric				
Catchment	Inputs	RMSE	MPRE	RB	NSE	KGE
Severn at Haw Bridge	7	63.0	41.5	−0.130	0.698	0.708
	28	48.1	38.6	−**0.014**	0.824	0.819
	7 + SM	42.6	**27.0**	−0.045	0.862	0.877
	7 + P	**42.3**	28.8	−0.033	**0.864**	**0.880**
Bedford Ouse at Roxton	7	8.34	56.5	−0.326	0.641	0.656
	28	7.00	40.4	−0.183	0.747	0.778
	7 + SM	5.86	**28.7**	−**0.127**	0.823	0.847
	7 + P	**5.82**	35.2	−0.150	**0.825**	**0.871**
Findhorn at Shenachie	7	12.3	69.1	−0.476	0.574	0.612
	28	12.4	63.7	−0.411	0.569	0.634
	7 + SM	**10.7**	**54.2**	−0.366	**0.678**	**0.690**
	7 + P	11.6	57.0	−**0.347**	0.625	0.650

taper off beyond 28 days and came at the expense of numerical stability during training. From the RB, all models had a tendency to underpredict, though the evidence for which underpredicts the least was inconclusive.

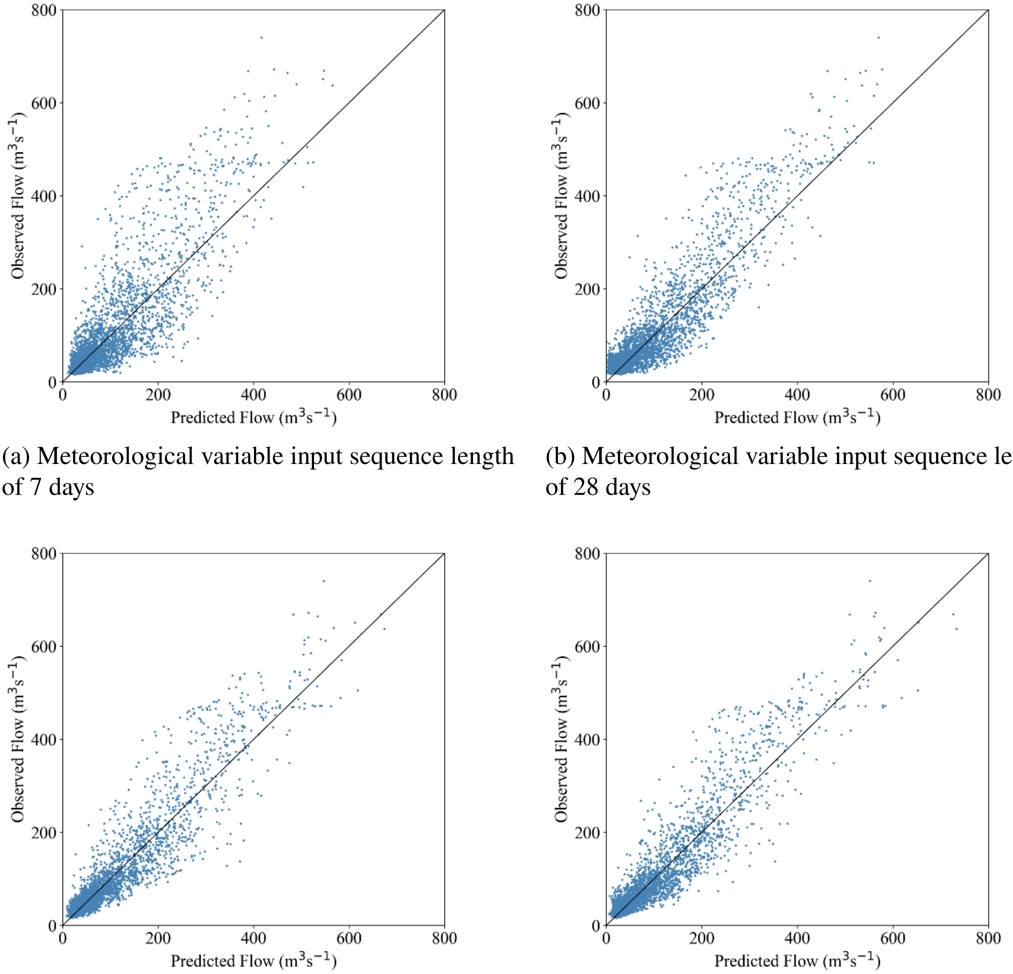
Through examination of the NSE and KGE, however, the overall performance for the Findhorn at Shenachie is clearly lower than that for the other catchments. This was expected due to the smaller size of the catchment potentially leading to a more rapid flow response. Compared to the Bedford Ouse at Roxton, the Findhorn at Shenachie is smaller yet the mean flow rate is higher. The inference is that the response from the catchment requires a finer temporal scale than the daily scale used here. Regardless, we can still see the same pattern emerging that the short variable record in conjunction with either soil moisture or the antecedent proxies results in a model with a high degree of skill. Furthermore, all models

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

exceed our benchmarks and the higher KGE relative to the NSE indicates that the models are predicting peaks and are representing the variability of the flow patterns well.

We present a subset of the results, those for the Severn at Haw Bridge: with the predictions against observations in Figure 3; exemplar streamflow time series observations and predictions for the year 2012 (from the test set) in Figure 4; and the percentage relative error for different percentile bands in Figure 5. The figures for the other catchments are shown in the Supplementary Materials.

Across all three catchments, baseflow and peak flow events are less well captured in the 28-day and 7-day models; whereas this is noticeable for winter peaks in the larger two catchments, all model implementations struggle with the summer peaks for the Findhorn at Shenachie. Using the soil moisture and proxy variable models result in a significant improvement in the representation of baseflow and a more modest improvement in the representation of peak flow events. Model results, while not perfect, are approaching a level of performance that would be considered robust. The set of catchments tested here is, obviously, small and further work may be required to refine the approach further. Having said that, with this framework in place, an alternative departure point from this work might be to move into the multiple



(a) Meteorological variable input sequence length of 7 days

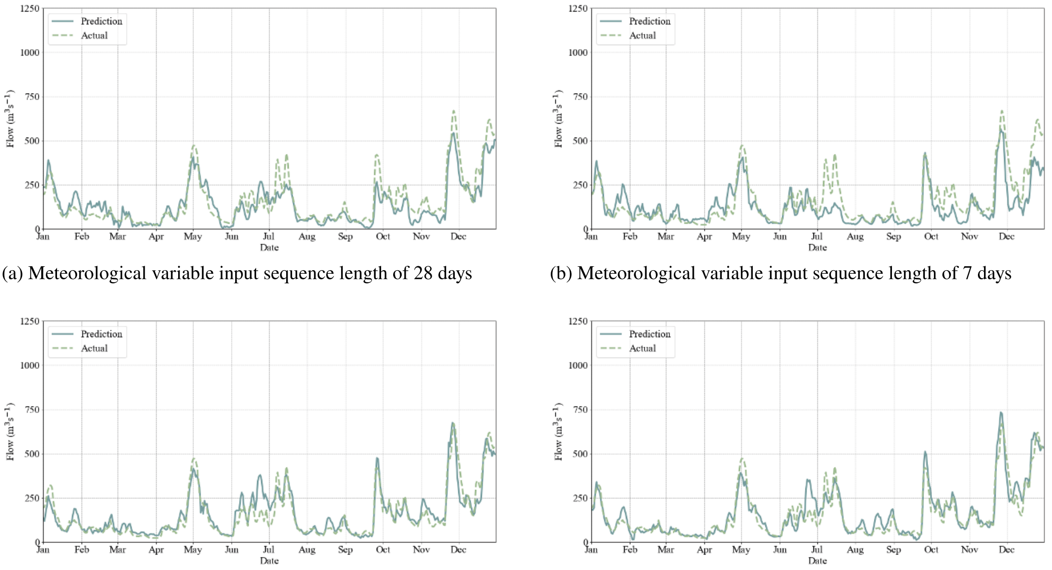
(b) Meteorological variable input sequence length of 28 days

(c) Meteorological variable input sequence length of 7 days with soil moisture variables

(d) Meteorological variable input sequence length of 7 days with antecedent proxy variables

Figure 3. *Predictions against observations for the Severn at Haw Bridge from the test set of predictions generated using different feature sets as inputs to the MLP model.*

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press



(a) Meteorological variable input sequence length of 28 days

(b) Meteorological variable input sequence length of 7 days

(c) Meteorological variable input sequence length of 7 days with soil moisture variables

(d) Meteorological variable input sequence length of 7 days with antecedent proxy variables

Figure 4. *Comparative streamflow time series for the Severn at Haw Bridge in the year, 2012, with both predictions and observations using different feature sets as inputs to the MLP model.*

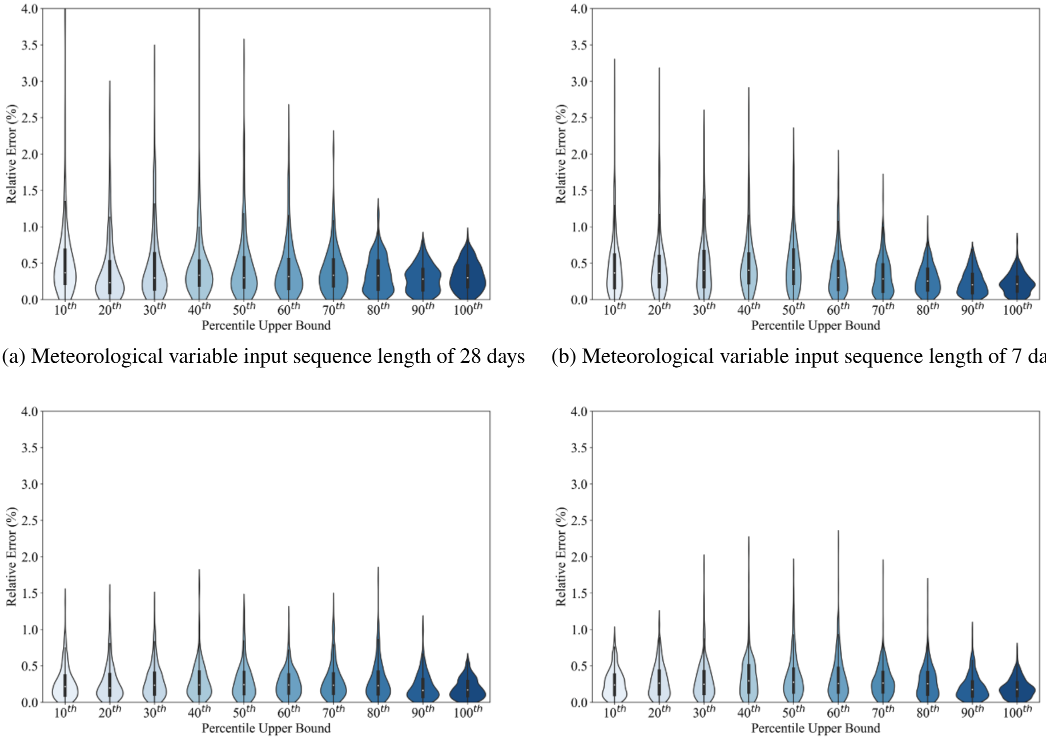
catchment setting to prove that this is a suitable climatic framework for constructing a generalizing model; in that context, the catchment descriptors would come into play and, perhaps with more data, enable the learning of small catchment response versus larger catchment response.

4.3. Network sensitivity

Using the input perturbation algorithm, we assess the sensitivity of the two separate networks trained on either the soil moisture variable set or the antecedent proxy variable set. Sensitivity has been calculated with perturbations in either direction, to account for the varying response to lower temperatures and higher precipitation when compared with lower precipitation and higher temperatures; arguably, utilization of the gradient method of sensitivity analysis might avoid this issue and will be explored in further work. A subset of results averaged across the three study catchments is presented in Table 3. Model sensitivity to the soil moisture variables and the antecedent proxy variables is of a similar order of magnitude; if we consider the aggregate sensitivity to these variable sets, then it is roughly equivalent over these two variable sets for the positive perturbation but higher for the antecedent variable set for the negative perturbation. Essentially, the response of the network to these variables appears to be similar. In terms of whether or not all the proxies or soil moisture levels are required to force the model, due to the varying information represented and the relative sensitivities, then it would be hard to argue that any one variable was not required in either the soil moisture to antecedent proxy variable sets, though soil moisture Level 1 might perhaps be well represented by and surplus to soil moisture Level 2 and short-term precipitation.

When examining the response due to the daily meteorological variables, it is worth noting that the Findhorn at Shenachie, as a smaller, less permeable and therefore flashier catchment, is far more sensitive to the most recent 24 hours worth of rainfall when compared to the larger catchment that feeds the Severn at Haw Bridge. Overall, the closer to the flow event, the more impactful the perturbation. The network sensitivity drops off substantially after 5 days (backward from the event), as shown in Figure 6, and sensitivity to daily meteorological forcing is superseded by sensitivity to the soil moisture levels or antecedent proxies after this point. Thus, the antecedent proxies are capable of replacing longer

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press



(a) Meteorological variable input sequence length of 28 days

(b) Meteorological variable input sequence length of 7 days

(c) Meteorological variable input sequence length of 7 days with soil moisture variables

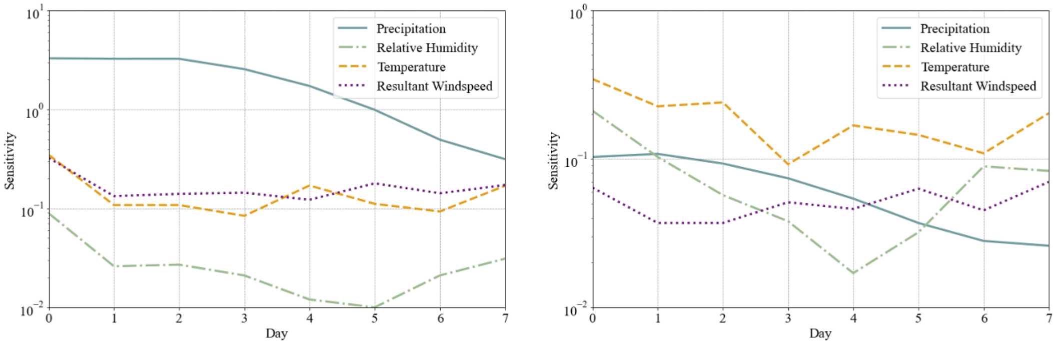
(d) Meteorological variable input sequence length of 7 days with antecedent proxy variables

Figure 5. *Violin plots of percentage relative error for each of the 10th percentile bands of flow magnitude (with the upper bound marked on the scale and the lower bound being the preceding upper bound to the left) between observations and predictions for the Severn at Haw Bridge using different feature sets as inputs to the MLP model.*

Table 3. *Average model sensitivity to individual soil moisture and antecedent proxy inputs using positive and negative perturbations for the two different model setups*

Variable	$\delta+$	$\delta-$
	Soil moisture model	
SM_{L1}	0.715	0.313
SM_{L2}	1.422	0.552
SM_{L3}	1.635	0.371
SM_{L4}	0.362	0.273
	Antecedent proxy model	
$\mu_{p:30}$	1.128	0.352
$\mu_{p:90}$	0.441	0.228
$\mu_{p:180}$	0.696	0.367
$\mu_{t:30}$	0.574	1.119
$\mu_{t:90}$	0.570	1.466
$\mu_{t:180}$	0.719	0.688

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press



(a) Positive perturbation

(b) Negative perturbation

Figure 6. *Average network sensitivity to daily meteorological variables under a positive perturbation (left) and negative perturbation (right), with logarithmic scales on the y axis.*

meteorological records and the soil moisture levels, likely encapsulating much of the information encoded within the aforementioned variables.

4.4. Study limitations

We note two areas in particular that were beyond the scope of this paper but are worth future exploration. The first being that, although we have quantified the error across percentile ranges, we have not developed the model to allow for prediction and reconstruction of the flow volume into base and peak flows. Though the output data we used were average daily flow, empirical methods exist for the reconstruction of base and peak flows that can be parameterized or trained on a per catchment basis (Fathzadeh et al., 2017; Jimeno-Sáez et al., 2017). The second area is on the quantification of uncertainty, in terms of both the aleatoric and epistemic uncertainty. Inadequate quantification of uncertainty can hinder the utilization of artificial intelligence approaches in decision and policy-making processes (Borrego et al., 2008; Cabaneros and Hughes, 2022) and provision of model uncertainty therefore aids adoption.

5. Conclusion

Through our approach, we have minimized the internal measured data burden by replacing relatively inaccessible variables with more accessible proxies. Subsequent model development can focus on more readily obtainable data in addition to offering a reduction in input space dimensionality that reduces the total number of model parameters. Our belief is that model parsimony helps improve model interpretability and could increase adoption of machine learning methods in operational hydrology. Leveraging domain knowledge, in this case through physically reasoned variable selection, can deliver high performance in conjunction with, rather than at the expense of, model parsimony.

While more modern neural network architectures may offer performance improvements and even uncertainty predictions (for example, Gaussian or Neural Processes (Garnelo et al., 2018; Rasmussen, 2004)), it was not our intention here to obtain best in class performance. Instead, we have highlighted how it is possible to limit the number of variable records required as input; and use a shortened temporal subset of forcing variables alongside climatological proxies to represent the longer-term hidden state of a catchment (here, the water storage as soil moisture). This in turn enables a more generic, yet still expressive, input space with which to force hydrological models that is easily generalized to other catchments.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2024.48.

Author contribution. Conceptualization: R.E.R., A.M.; Data curation: R.E.R.; Funding acquisition: R.E.R., S.H., A.M., E.S.; Methodology: R.E.R., A.M., D.K.; Project administration: R.E.R., S.H., A.M., E.S.; Software: R.E.R., S.H.; Supervision: S.H., A.M., E.S.; Visualization: R.E.R., A.M.; Writing—Original draft: R.E.R., D.K.; Writing—Review and editing: R.E.R., D.K.

Competing interest. The authors declare no competing interests exist.

Data availability statement. The data required to reproduce these results are available at the NRFA (Dixon et al., 2013) using the data portal https://nrfa.ceh.ac.uk/data/search or API https://nrfaapps.ceh.ac.uk/nrfa/nrfa-api.html and from the Copernicus Climate Data Store https://doi.org/10.24381/cds.143582cf. We have made the code to run the model available at https://doi.org/10.5281/zenodo.13861688.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. This research was supported by a grant from the Engineering and Physical Sciences Research Council (EP/R512461/1). This research was supported by a Fellowship from the Royal Commission for the Exhibition of 1851.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2024.48.

References

Abbott MB, Bathurst JC, Cunge JA, O'Connell PE and Rasmussen J (1986a) An introduction to the European hydrological system — Systeme Hydrologique Europeen, "SHE," 1: History and philosophy of a physically-based, distributed modelling system. *Journal of Hydrology 87*(1–2), 45–59.

Abbott MB, Bathurst JC, Cunge JA, O'Connell PE and Rasmussen J (1986b) An introduction to the European hydrological system — Systeme Hydrologique Europeen, "SHE," 2: Structure of a physically-based, distributed modelling system. *Journal of Hydrology 87*(1–2), 61–77.

Abrahart RJ, Anctil F, Coulibaly P, Dawson CW, Mount NJ, See LM, Shamseldin AY, Solomatine DP, Toth E and Wilby RL (2012) Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Progress in Physical Geography: Earth and Environment 36*(4), 480–513.

Aichouri I, Hani A, Bougherira N, Djabri L, Chaffai H and Lallahem S (2015) River flow model using artificial neural networks. *Energy Procedia 74*, 1007–1014.

Ali S and Shahbaz M (2020) Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. *Modeling Earth Systems and Environment 6*(3), 1645–1656.

Arnold JG, Srinivasan R, Muttiah RS and Williams JR (1998) Large area hydrologic MODELING and assessment part i: Model development. *Journal of the American Water Resources Association 34*(1), 73–89.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000) Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering 5*(2), 115–123.

Ayzel G and Heistermann M (2021) The effect of calibration data length on the performance of a conceptual hydrological model versus LSTM and GRU: A case study for six basins from the CAMELS dataset. *Computers & Geosciences 149*, 104708.

Balog I, Caputo G, Iatauro D, Signoretti P and Spinelli F (2023) Downscaling of hourly climate data for the assessment of building energy performance. *Sustainability 15*(3), 2762.

Bartens A and Haberlandt U (2021) Flood frequency analysis using mean daily flows vs. instantaneous peak flows. *Hydrology and Earth System Sciences Discussions 2021*, 1–25.

Bishop CM (2016) *Pattern Recognition and Machine Learning.* Information Science and Statistics. New York, NY: Springer New York; softcover reprint of the original 1st edition 2006 (corrected at 8th printing 2009) edition.

Borrego C, Monteiro A, Ferreira J, Miranda AI, Costa AM, Carvalho AC and Lopes M (2008) Procedures for estimation of modelling uncertainty in air quality assessment. *Environment International 34*(5), 613–620.

Cabaneros SM and Hughes B (2022) Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting. *Environmental Modelling & Software 158*, 105529.

Calver A, Stewart E and Goodsell G (2009) Comparative analysis of statistical and catchment modelling approaches to river flood frequency estimation: River flood frequency estimation. *Journal of Flood Risk Management 2*(1), 24–31.

Cao M, Alkayem NF, Pan L and Novák D (2016) Advanced methods in neural networks-based sensitivity analysis with their applications in civil engineering. In Rosa JLG (ed.), *Artificial Neural Networks - Models and Applications*. InTech.

Cao M and Qiao P (2008) Neural network committee-based sensitivity analysis strategy for geotechnical engineering problems. *Neural Computing and Applications 17*(5–6), 509–519.

Dawson CW, Abrahart RJ, Shamseldin AY and Wilby RL (2006) Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology 319*(1–4), 391–409.

De Boor C (2001) *A Practical Guide to Splines. Number V. 27 in Applied Mathematical Sciences*, revised edition. New York: Springer.

Department for Environment, Food & Rural Affairs (2024) Structure of the agricultural industry in England and the UK at June. https://www.gov.uk/government/statistical-data-sets/structure-of-the-agricultural-industry-in-england-and-the-uk-at-june

Department for Levelling Up, Housing and Communities (2023) Land use change statistics. https://www.gov.uk/government/collections/land-use-change-statistics10.5194/hess-2019-327

Devi GK, **Ganasri BP and Dwarakish GS** (2015) A review on hydrological models. *Aquatic Procedia 4*, 1001–1007.

Ding J, **Haberlandt U and Dietrich J** (2015) Estimation of the instantaneous peak flow from maximum daily flow: A comparison of three methods. *Hydrology Research 46*(5), 671–688.

Dixon H, **Hannaford J and Fry MJ** (2013) The effective management of national hydrometric data: Experiences from the United Kingdom. *Hydrological Sciences Journal 58*(7), 1383–1399.

Dong X, **Lai X**, **Wang Y**, **Dong W**, **Zhu J**, **Dong L and Cen S** (2022) Applicability evaluation of multiple sets of soil moisture data on the tibetan plateau. *Frontiers in Earth Science 10*, 872413.

Fathzadeh A, **Jaydari A and Taghizadeh-Mehrjardi R** (2017) Comparison of different methods for reconstruction of instantaneous peak flow data. *Intelligent Automation & Soft Computing 23*(1), 41–49.

Faulkner DS, **Francis O and Lamb R** (2012) Greenfield run off and flood estimation on small catchments: Small catchments' greenfield run off and flood estimation. *Journal of Flood Risk Management 5*(1), 81–90.

Fill HD and Steiner AA (2003) Estimating instantaneous peak flow from mean daily flow data. *Journal of Hydrologic Engineering 8*(6), 365–369.

Fuller WE (1914) Flood flows. *Transactions of the American Society of Civil Engineers 77*(1), 564–617.

Gao S, **Huang Y**, **Zhang S**, **Han J**, **Wang G**, **Zhang M and Lin Q** (2020) Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology 589*, 125188.

Garnelo M, **Rosenbaum D**, **Maddison C**, **Ramalho T**, **Saxton D**, **Shanahan M**, **Teh YW**, **Rezende D and Eslami SMA** (2018) Conditional neural processes. In Dy J and Krause A (eds.), *Proceedings of the 35th International Conference on Machine Learning, Volume 80 of Proceedings of Machine Learning Research*. PMLR, pp. 1704–1713.

Gevrey M, **Dimopoulos I and Lek S** (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling 160*(3), 249–264.

Govindaraju RS, **Ramachandra Rao A and Singh VP** (eds.) (2000) *Artificial Neural Networks in Hydrology, Volume 36 of Water Science and Technology Library*. Netherlands, Dordrecht: Springer.

Gray DM (ed.) (1973) *Handbook on the Principles of Hydrology: With Special Emphasis Directed to Canadian Conditions in the Discussions, Applications, and Presentation of Data*. Port Washington, N.Y: Water Information Center, Inc.

Gupta HV, **Kling H**, **Yilmaz KK and Martinez GF** (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology 377*(1–2), 80–91.

Hersbach H, **Bell B**, **Berrisford P**, **Biavati G**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Rozum I**, **Schepers D**, **Simmons A**, **Soci C**, **Dee D and Thépaut J-N** (2018) ERA5 hourly data on single levels from 1979 to present. https://doi.org/10.24381/cds.adbb2d47.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S and Thépaut J-N** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049.

Institute of Hydrology (2008) *Flood Estimation Handbook. Number 2 in Flood Estimation Handbook / Institute of Hydrology.* Wallingford: Centre for Ecology and Hydrology.

Jaiswal RK, **Ali S and Bharti B** (2020) Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science 10*(1), 48.

Jimeno-Sáez P, **Senent-Aparicio J**, **Pérez-Sánchez J**, **Pulido-Velazquez D and Cecilia J** (2017) Estimation of instantaneous peak flow using machine-learning models and empirical formula in peninsular Spain. *Water 9*(5), 347.

Kidd C, **Becker A**, **Huffman GJ**, **Muller CL**, **Joe P**, **Skofronick-Jackson G and Kirschbaum DB** (2017) So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society 98*(1), 69–78.

Kingma DP and Ba J (2017) Adam: A method for stochastic optimization. *arXiv*; arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980.

Krabbenhoft CA, **Allen GH**, **Lin P**, **Godsey SE**, **Allen DC**, **Burrows RM**, **DelVecchia AG**, **Fritz KM**, **Shanafield M**, **Burgin AJ**, **Zimmer MA**, **Datry T**, **Dodds WK**, **Jones CN**, **Mims MC**, **Franklin C**, **Hammond JC**, **Zipper S**, **Ward AS**, **Costigan KH**, **Beck HE and Olden JD** (2022) Assessing placement bias of the global river gauge network. *Nature Sustainability 5*, 586–592.

Kumar DN, **Raju KS and Sathish T** (2004) River flow forecasting using recurrent neural networks. *Water Resources Management 18*(2), 143–161.

Lal P, **Singh G**, **Das NN**, **Colliander A and Entekhabi D** (2022) Assessment of ERA5-land volumetric soil water layer product using in situ and SMAP soil moisture observations. *IEEE Geoscience and Remote Sensing Letters 19*, 1–5.

Li M, **Wu P and Ma Z** (2020) A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *International Journal of Climatology 40*(13), 5744–5766.

Moriasi DN, **Arnold JG**, **Van Liew MW**, **Bingner RL**, **Harmel RD and Veith TL** (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE 50*(3), 885–900.

https://doi.org/10.1017/eds.2024.48 Published online by Cambridge University Press

Morris DG and Flavin RW (1990) A digital terrain model for hydrology. In *Proceedings of the 4th International Symposium on Spatial Data Handling*, International Geographical Union IGU, Commission on Geographic Information Systems, Department of Geography, The Ohio State University, Vol. *1*, pp. 250–262.

Morris DG and Flavin RW (1994) *Subset of UK 50 m by 50 m Hydrological Digital Terrain Model Grids. Technical Report*. Wallingford: NERC, Institute of Hydrology.

Nash JE and Sutcliffe JV (1970) River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology 10*(3), 282–290.

Nogueira M (2020) Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences. *Journal of Hydrology 583*, 124632.

Pizarroso J, **Portela J and Muñoz A** (2022) NeuralSens: Sensitivity analysis of neural networks. *Journal of Statistical Software 102*(7), 1–36.

Ramachandran P, **Zoph B and Le QV** (2017) Searching for activation functions. *arXiv*; arXiv: 1710.05941. https://doi.org/10.48550/arXiv.1710.05941.

Rasmussen CE (2004) Gaussian processes in machine learning. In Bousquet O, von Luxburg U and Rätsch G (eds.), *Advanced Lectures on Machine Learning*, Vol. *3176*. Berlin Heidelberg, Berlin, Heidelberg: Springer, pp. 63–71.

Refsgaard JC, **Storm B and Clausen T** (2010) Système Hydrologique Europeén (SHE): Review and perspectives after 30 years development in distributed physically-based hydrological modelling. *Hydrology Research 41*(5), 355–377.

Robson A and Reed D (2008) *Statistical Procedures for Flood Frequency Estimation. Number 3 in Flood Estimation Handbook / Institute of Hydrology*. Wallingford: Centre for Ecology and Hydrology.

Rogelis MC, **Werner M**, **Obregón N, and Wright N** (2016) Hydrological model assessment for flood early warning in a tropical high mountain basin. Available at https://hess.copernicus.org/preprints/hess-2016-30/.

Ruder S (2017) An overview of gradient descent optimization algorithms. *arXiv*; arXiv:1609.04747. https://doi.org/10.48550/arXiv.1609.04747.

Rumelhart DE, **Hinton GE and Williams RJ** (1986) Learning representations by back-propagating errors. *Nature 323*(6088), 533–536.

Samuels P, **Huntington S**, **Allsop W and Harrop J** (2008) *Flood Risk Management: Research and Practice: Extended Abstracts Volume (332 Pages) + Full Paper CD-ROM (1772 Pages)*. CRC Press.

Srinivasan R, **Ramanarayanan TS**, **Arnold JG and Bednarz ST** (1998) Large area hydrologic MODELING and assessment part ii: Model application. *Journal of the American Water Resources Association 34*(1), 91–101.

Tabor K and Williams JW (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications 20*(2), 554–565.

Tarek M, **Brissette FP and Arsenault R** (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences 24*(5), 2527–2544.

UK Centre for Ecology & Hydrology (2022) UK National River Flow Archive Data. https://nrfa.ceh.ac.uk/data.

Wilby RL and Quinn NW (2013) Reconstructing multi-decadal variations in fluvial flood risk using atmospheric circulation patterns. *Journal of Hydrology 487*, 109–121.

Wouter J, **Knoben M**, **Freer JE and Woods RA** (2019) Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Available at https://hess.copernicus.org/articles/23/4323/2019/.

Yilmaz KK, **Gupta HV and Wagener T** (2008) A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research 44*(9). https://doi.org/10.1029/2007WR006716.

Cite this article: Rouse RE, Khamis D, Hosking S, McRobie A and Shuckburgh E (2025). Streamflow prediction using artificial neural networks and soil moisture proxies. *Environmental Data Science*, 4: e5. doi:10.1017/eds.2024.48