Journal of Environmental Management 354 (2024) 120404

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Coupling machine learning and physical modelling for predicting runoff at catchment scale

Sergio Zubelzu [a,*], Abdulmomen Ghalkha [b], Chaouki Ben Issaid [b], Andrea Zanella [c], Medhi Bennis [b]

[a] Departamento de Ingeniería Agroforestal, Universidad Politécnica de Madrid, Madrid, Spain
[b] Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
[c] Department of Information Engineering, University of Padova, Padova, Italy

ARTICLE INFO

Handling Editor: Jason Michael Evans

Keywords:
Hydrology
Machine learning
Physical modelling
Runoff
Catchment

ABSTRACT

In this paper, we present an approach that combines data-driven and physical modelling for predicting the runoff occurrence and volume at catchment scale. With that aim, we first estimated the runoff volume from recorded storms aided by the Green-Ampt infiltration model. Then, we used machine learning algorithms, namely LightGBM (LGBM) and Deep Neural Network (DNN), to predict the outputs of the physical model fed on a set of atmospheric variables (relative humidity, temperature, atmospheric pressure, and wind velocity) collected before or immediately after the beginning of the storm. Results for a small urban catchment in Madrid show DNN performed better in predicting the runoff occurrence and volume. Moreover, enriching the input primary atmospheric variables with auxiliary variables (e.g., storm intensity data recorded during the first hour, or rain volume and intensity estimates obtained from auxiliary regression methods) largely increased the model performance. We show in this manuscript data-driven algorithms shaped by physical criteria can be successfully generated by allowing the data-driven algorithm learn from the output of physical models. It represents a novel approach for physics-informed data-driven algorithms shifting from common practices in hydrological modelling through machine learning.

1. Introduction

In 2017, a set of 230 experts, including well known hydrologists and scientists from other related disciplines, highlighted several unsolved questions in hydrology (Blösch et al., 2017). They arose some issues related to hydrologic laws and their suitability at different scales, the use of historical data *vs* soft data, the reduction of the amount of model structural/parameter/input uncertainty in hydrological prediction, among others. Following the conclusions presented in that work, the comprehensive understanding of the complex interactions among the physical processes underlying the hydrological systems still remained elusive.

The accurate modelling of hydrological systems is complex, involving highly variable and interlinked distant processes, such as precipitation, infiltration and flood routing. Theories for modelling such phenomena were proposed long time ago (see, for example, Thornthwaite and Holzman, 1939; Darcy, 1856; Philip, 1957; Richards, 1931; Saint Venant, 1871). Nonetheless, physical models still suffer from major problems, such as mathematical complexity, unreliability of parameterization when modelling large and heterogeneous watersheds, inability to adequately handle extreme spatial and temporal variability, or lack of random criteria in modelling the evolution of hydrological systems. Physical models are hence rigorous from a conceptual perspective but often prohibitively complex when large-scale accurate results are sought.

Given the performance achieved by data-driven models in different fields, and the always larger amount of data available, many hydrologists have then explored in the last years the suitability of applying machine-learning (ML) for modelling hydrologic processes. Some recent and comprehensive overviews of this body of work are provided in Mohammadi (2022), Mosaffa et al. (2022) or Zounemat-Kermani et al. (2021).

By analysing the literature, we can observe that the application of ML algorithms for modelling hydrological processes base on either

* Corresponding author.
E-mail addresses: sergio.zubelzu@upm.es (S. Zubelzu), Abdulmomen.Ghalkha@oulu.fi (A. Ghalkha), chaouki.benissaid@oulu.fi (C. Ben Issaid), andrea.zanella@unipd.it (A. Zanella), mehdi.bennis@oulu.fi (M. Bennis).

https://doi.org/10.1016/j.jenvman.2024.120404
Received 23 September 2023; Received in revised form 29 January 2024; Accepted 13 February 2024
Available online 19 February 2024

autoregressive approaches (Tikhamarine et al., 2020; Zuo et al., 2020; Shabri and Suhartono, 2012) or physical causality (Bui et al., 2020; Wang et al., 2020; Pourghasemi et al., 2020; Dariane and Azimi, 2018; Kratzert et al., 2019; Alizadeh et al., 2018; Zhong et al., 2023; Yokoo et al., 2022; Bhasme et al., 2022). The first group comprises models where outputs and inputs are time-lagged variables. Within the second group we can find models that seek to bind the model's inputs and outputs by physical causality. This approach explores the so-called physics informed ML pathway for modelling hydrological processes which consists in shaping the machine learning algorithm following hydrological principia.

However, despite the growing interest dedicated to these approaches and their apparent good performance, concerns raise when data-driven algorithms are analysed in detail (Zanella et al., 2023). For what regards to autoregressive approaches, ML algorithms must ensure the temporal dimension of the input data they use is correct. Rainfall-runoff hydrology is triggered by asynchronous events (storms) and using fixed-period time series (monthly, daily or even hourly), as is mostly done in previous published works, is not suitable for properly observing these processes. The focus has to be on hyetographs rather than the cumulative precipitation in long-time intervals. Similarly, the time lags considered for feeding the model must be coherent and going back several weeks or months is only defensible if the catchment physical and territorial characteristics allow the observed hydrological variable to be dependent on what happened weeks or months ago. It is especially questionable for some previous published works that found good performance in models predicting the streamflow fed on either 7 or 8 monthly lagged precipitation or streamflow records.

We strongly believe there is also still large field for improving the way physics informed ML algorithms are being built. The proper selection of physically related inputs and outputs is the key point we argue previous works are not addressing correctly. Previous attempts exploit existing datasets to build predictive models on poorly connected variables with no structured causal relationship defined by theoretical models. Most of the existing works select a set of weather, territorial or geographical parameters to feed the algorithms in the search for the combination yielding the best performance in predicting the output while obviating any physical theory supporting the relationship between input and output.

In this paper, we seek a novel approach for designing data-driven algorithms based on physical theories that will help overcome the presented issues. We move away from previous ideas by interlocking both physically-based and data-driven modelling approaches, so as to build ML algorithms that embody the causal relationships defined by physical theories.

We applied this idea to develop a data-driven algorithm for predicting the occurrence and volume of runoff from atmospheric variables recorded before the storm starts. We acted as follows: 1) we first extracted storm hyetographs from precipitation records; 2) then used the Green-Ampt theoretical model to determine whether each storm generates runoff and, in case, of what volume; 3) we trained some ML algorithms with atmospheric measurements (temperature, wind velocity, relative humidity and atmospheric pressure) collected just before the beginning of each storm event with the intent of predicting the occurrence and volume of runoff estimated by the Green-Ampt model.

2. Materials and methods

We retrieved weather records of precipitation (P), relative humidity (HR), temperature (T), atmospheric pressure (PB) and wind velocity (VV) from six weather stations[1] located in Madrid city (Spain). We used hourly data from January 2019 to October 2022 (in total, 33,576 hourly samples of each variable). After cleaning spurious records, we extracted the storm hyetographs of each rainstorm event (defined as any set of positive precipitation records between two zero precipitation values). Then, we determined the aggregated volume (V), duration (D), average rainfall intensity (Iav) and maximum rainfall intensity (Im) of each storm event. We also stored HR, T, PB and VV records one, two, three, four and 5 h before the storm beginning and during the storm event as well. We then applied linear interpolation to increase the time granularity of the measurements to 10 min.

These hyetographs were then used as input variables to the Green-Ampt model as presented in Chow et al. (1988), from which we estimated the occurrence of runoff and its characteristics, namely: volume (VR), maximum intensity (Rim), and average intensity (RIav).

Our comprehensive assumption is that there exists physical causality between the evolution of the atmospheric variables (HR, T, PB, VV) before the storm starts, the derived storms characteristics and, consequently, the runoff occurrence and volume estimated by the Green-Ampt model. Under these assumptions, we fed the data-driven algorithm on primary atmospheric variables (input) potentially explaining the final runoff characteristics (output), as sketched in see Fig. 1. The physical model in this case provide the ML algorithm with the output to be forecasted while the inputs are the set of atmospheric records collected before or immediately after the beginning of the storm.

2.1. Physical modeling: the Green-Ampt model

Let i(t) denote the rainfall intensity and f(t) be the potential infiltration rate of the ground at time t. Moreover, let P(t) and F(t) be the aggregated precipitation and infiltration values, respectively, up to time t.

Then, as long as i(t) < f(t), rainfall completely infiltrates the ground, and runoff does not occur. In this condition, we have P(t) = F(t). On the contrary, if i(t) > f(t), the ground is no longer capable of absorbing all rainfall and the excess will flow on surface, creating runoff. The volume of runoff up to time t, when it occurs, is then given by VR(t) = P(t) − F(t).

Following Chow et al. (1988), the Green-Ampt model provides the potential infiltration rate f(t) as

$$f(t) \cong k_s\left(1+\frac{\tau_f\cdot\Delta\theta}{F(t)}\right) \tag{1}$$

and

$$F(t)=k_s\cdot t+\tau_f\cdot\Delta\theta\cdot Ln\left(1+\frac{F(t)}{\tau_f\cdot\Delta\theta}\right) \tag{2}$$

where ks, Δθ and τ_f stand for the saturated hydraulic conductivity, the difference between the saturated and the initial soil water content and the wetting front suction head, respectively. This last parameter, in turn, can be estimated following Neuman (1976).

$$\tau_f=\frac{1}{K_s}\int_{h(\theta_i)}^{h(\theta_s)} K(\theta)dh(\theta) \tag{3}$$

aided by Van Genuchten (1980) and Mualem's (1976) models for the soil conductivity and water retention curves:

$$\frac{\theta-\theta_r}{\theta_s-\theta_r}=\theta^*=[1+(\alpha h(\theta))^n]^{-m}; \tag{4}$$

$$k(\theta)=k_s\theta^{*0.5}\left[1+\left(1-\theta^{*\frac{1}{m}}\right)^m\right]^2. \tag{5}$$

in the equations above, the parameters θs and θr indicate the saturated and residual soil moisture levels, θ* the normalised soil moisture and m, n and α a set of soil-type dependent parameters.

We used the previous approach for estimating the occurrence of runoff (binary variable taking value 1 in case runoff occurs at any time

[1] "J.M.D. Moratalaz", "J.M.D. Villaverde", "Centro Municipal de Acústica", "J.M.D. Hortaleza", "Peñagrande" and "Plaza Elíptica".

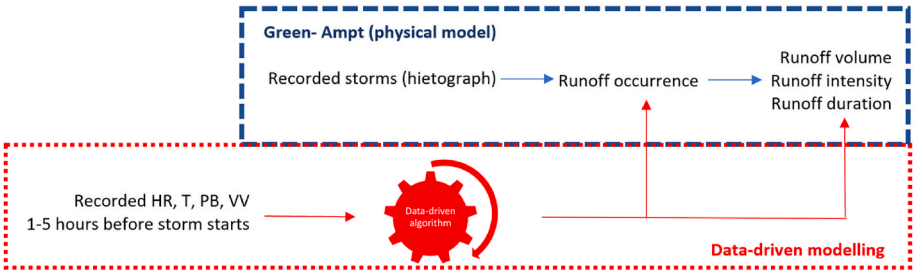

Fig. 1. Proposed conceptual approach for coupling data-driven and physical modelling.

during the storm and 0 otherwise) and, if positive, the volume of runoff (VR(RD), where RD is the overall runoff duration), the maximum intensity RIm = $\max_t$\{VR(t)-VR(t-1)\}, and the average intensity RIav = VR (RD)/RD of each storm event.

2.2. Data-driven inference algorithms

Let {Xi, yi} represent a generic dataset, where Xi = {x_1, x_2, …, x_n} are the i-th input variable vector and yi is the corresponding i-th output of the system. The goal of data-driven inference algorithms is to estimate a function Z(x) that minimizes some loss function L(y, Z(x)):

$$\widehat{Z} = \arg\min_Z E_{y,x}[L(y, Z(x))]. \tag{6}$$

in our study, we used LGBM and DNN algorithms. The mathematical basis of both are given below.

1) LGBM is a gradient boosting framework that uses a tree-based learning algorithm. Gradient boosting (GBDT) is a way to solve (6) in which Z(x) is represented as an ensemble of weak predictors, which are decision trees. The GBDT model with N stages can be expressed as:

$$Z(x) = \sum_{j=1}^{N} h_j(x;\theta_j) \tag{7}$$

where $h_j(x;\ \theta_j)$ is the j-th weak predictor, and θ_j is its parameter. Therefore, the model at the iteration *j* can be formulated as:

$$Z_j(x) = Z_{j-1}(x) + h_j(x;\theta_j). \tag{8}$$

Therefore, GBDT avoids the high complexity encountered when trying to solve (6) with single-step optimization by dividing it into smaller sub-problems. In addition, LGBM uses a histogram-based approach to calculate the optimal splitting points for each feature. This involves dividing the feature values into a small number of discrete bins and then finding the optimal split point based on the histograms of these bins. This approach reduces the computational cost of determining the optimal splitting points and allows for faster training.

2) DNNs are well suited for data with complex patterns and relationships difficult to be modelled using traditional ML algorithms. DNNs consist of multiple layers of interconnected neurons, with each layer transforming the input data in a way that allows the next layer to extract more complex features from it. Each layer of the network consists of multiple neurons that perform a linear transformation followed by a non-linear activation function. By stacking many layers together, DNNs can learn highly abstract and hierarchical representations of data, which make them well suited for a wide range of machine learning tasks.

The mathematical representation of a single neuron in a layer can be written as:

$$z_j^{(l)} = \sum_{i=1}^{n^{(l-1)}} \omega_{ij}^{(l)} x_i^{(l-1)} + b_j^{(l)} \tag{9}$$

where *l* denotes the layer index, *j* the neuron index in that layer, $n^{(l-1)}$ the number of neurons in the previous layer, $w_{ij}^{(l)}$ the weight of the connection between neuron *i* in layer *(l − 1)* and neuron *j* in layer *l*, $x_i^{(l-1)}$ the output of neuron *i* in layer *(l − 1)*, $b_j^{(l)}$ the bias of neuron *j* in layer *l*, and $z_j^{(l)}$ the weighted sum of the inputs to neuron *j* in layer *l*. The output of a single neuron in the layer *l* can be written as:

$$a_j^{(l)} = f^{(j)}\left(z_j^{(l)}\right), \tag{10}$$

where $f^{(j)}(.)$ denotes the activation function at the *j*-th layer. The output of each layer is then passed as input to the next layer until we reach the last output layer, where the final output of the network is computed as:

$$y = f^{(L)}\left(z^{(L)}\right) \tag{11}$$

being z(L) the weighted sum of the inputs to the output layer, and $f^{(L)}$(.) the activation function of the output layer.

The entire DNN can be written as a composition of functions as:

$$y = f^{(L)}\left(f^{(L-1)}\left(\ldots f^{(2)}\left(f^{(1)}(x)\right)\right)\right). \tag{12}$$

The weights and biases of the neural network are learned through a process called backpropagation, where the error between predicted and true output is backpropagated through the network to adjust the weights and biases using gradient descent.

2.3. Proposed architectures of machine learning algorithms mimicking physical behaviour

We targeted both runoff occurrence (classification) and runoff characteristics (prediction) seeking to mimic the physical causality following different strategies as presented in the sections below and graphically illustrated in Fig. 2.

2.3.1. Runoff occurrence

To predict the occurrence of runoff, we tested different approaches, with progressively richer information sets, as explained below and displayed in Fig. 2.

(1.a) The first basic approach consists in applying both LGBM and DNN classifiers to predict the occurrence of runoff (output) directly from primary atmospheric variables (HR, T, PB, VV) collected from one to 5 h **before** the beginning of a storm.

(1.b) The second approach enriches the previous one by including in the input variables of the LGBM and DNN algorithms also the estimate of the aggregate rainfall volume and maximum rain intensity as given by a regression model from the atmospheric primary variables. Notice that this method still requires measurements collected **before** the beginning of a storm but, conversely to the previous one, explicitly estimates the intermediate variables that are known to drive the runoff process.

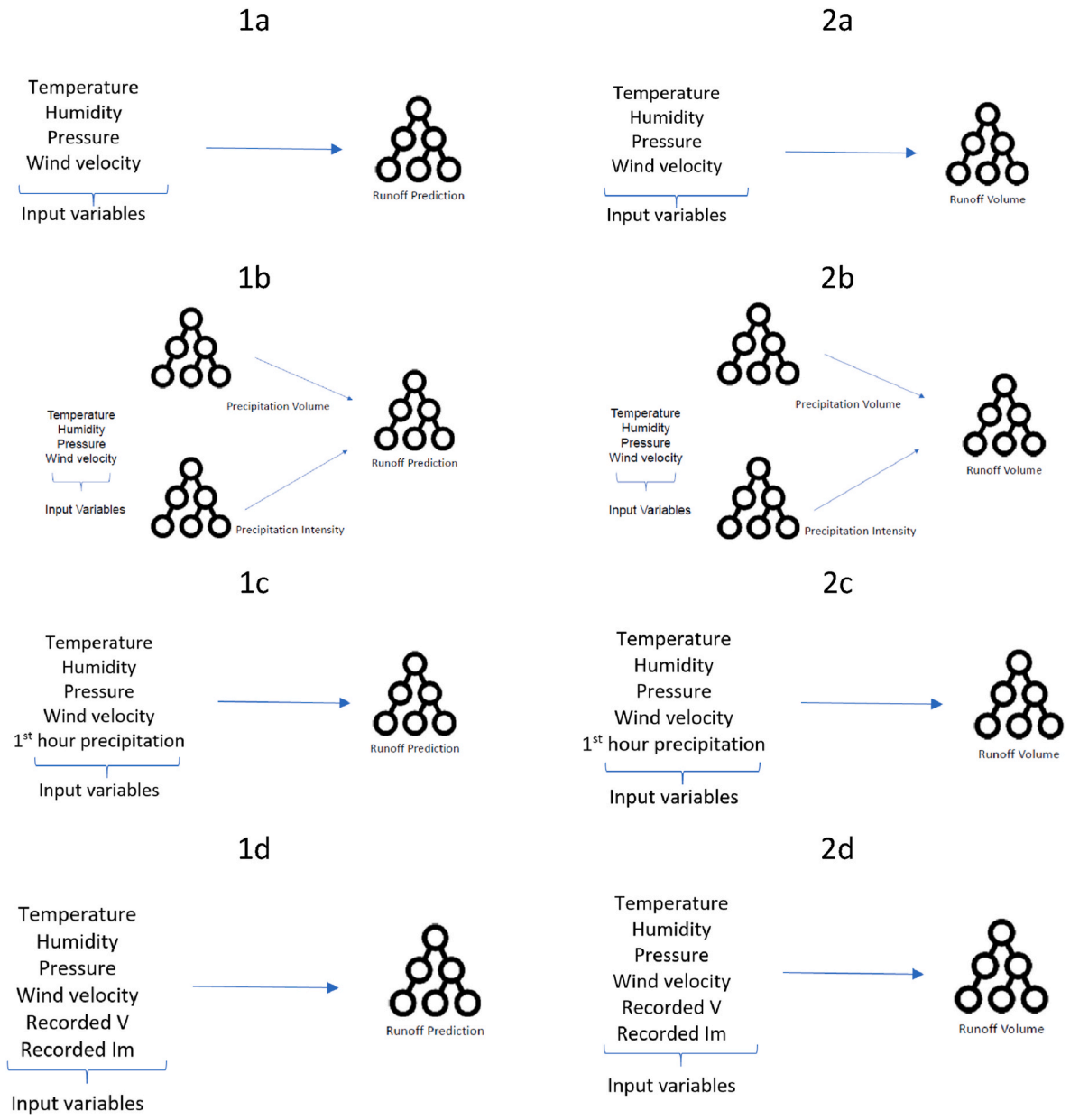

Fig. 2. Illustration of the models' construction with the predicted output from each step.

(1.c) The third approach replicates 1.a but incorporating the **recorded** precipitation volume during the first hour of the storm. In this case, the model can be applied only after the storm has started.
(1.d) The last approach is as 1.c but it uses the storm precipitation volume and maximum intensity recorded **after** the storm rather than those predicted with regression models. This model is used as benchmark, since it can only be applied after the end of the storm.

2.3.2. *Runoff volume prediction*

We applied a similar approach to predict the (estimated) volume of runoff:

(2.a) We employ LGBM and DNN regressors to predict the runoff volume using the primary atmospheric variables.
(2.b) Model 2.b enriches 2.a with the predicted values of the runoff intensity and volume, obtained through a regression model from the primary variables.
(2.c) Here we repeat 2.a, but using the actual measured precipitation volume during the first hour of the storm as auxiliary input variables in place of those estimated with the regression model.
(2.d) Finally, similarly to what done for the classification, we replicate 2.c by replacing the estimated precipitation volume and intensity with those measured at the end of the storm.

For running the models, data was divided into training and testing with ratios of 0.7 and 0.3 respectively.

2.4. *Performance metrics*

The performance of the runoff classifier is measured using both the F_1 score given by

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{13}$$

where TP, FP and FN stand for the true positive, false positive and false

negative numbers. The F_1 score combines precision (proportion of truly positive events over the total number of examples classified as positive) and recall (number of true positive examples over the total number of actual positive examples, i.e., the sum of the number of true positives and false negatives) metrics. In addition, we also consider a metric denoted by AUC, i.e., the area under the Receiver Operating Characteristic curve (ROC) that captures the trade-off between detecting true positives and false positives. Both metrics range from 0 (null) to 1 (outstanding).

For analysing the accuracy of the prediction model, we used the mean square error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (14)$$

where y_i and $\hat{y}$ represent the recorded and estimated values, respectively.

2.5. Case study

We have selected a 300.000 m^2 (non-urbanised and uncultivated) catchment located in Madrid city (within the so-called La Solana de Valdebebas urban development). A seasonal water course drains the runoff from North-West to South-East. The catchment is completely surrounded by urbanised areas (see Fig. 3).

We took four soil samples and estimated a loamy textural class from laboratory experiments (determining fractions of soil, sand and lime with soil densitometry and thus finding the textural classes aided by the USDA soil textural triangle). We gathered the soil physical parameters of a loamy soil required for running Neuman (1976), Van Genuchten (1980) and Mualem's (1976) models from Carsel and Parrish (1998).

3. Results

In this section we report the results of our analysis. Successively, we analyse the empirical and model-generated data using standard statistical tools with the aim of identifying possible relations that can be exploited to predict the runoff occurrence and characteristics from the observable variables. We finally consider the results given by the ML algorithms described in the previous section.

3.1. Storms characteristics and simulated runoff

In this section we present some descriptive statistics of the storms and explore the correlations between the storm and runoff characteristics (outputs) and the primary atmospheric variables (inputs).

3.1.1. Descriptive statistics

Paying attention to the storm hyetographs (see supplementary documentation), we can observe that only a fraction of the storm events produces runoff (8.5% on average across the year), which implies that the dataset is highly unbalanced for our purposes. As we can perceive from Table 1, the central tendency metrics indicate that storms last, on average, 1.66 h (50% last up to 1 h) delivering 2.3 mm of water (50% of them actually discharge 0.4 mm or less), with 0.6 mm/h and 13.54 mm/h of average and maximum rainfall intensities, respectively. Recorded data show strong positive skewness and kurtosis.

Table 1
Descriptive statistics. SD: standard deviation, Qn: n-th quartile, Skw: skew, K: Kurtosis. (all): all storms, (roff): storms producing runoff.

Variable	Maximum	Mean	Median	SD	Q_{25}/ Q_{75}	Skw	K
D (h) (all)	29	1.66	1	2.96	0.1/2	3.15	12.99
D (h) (roff)	29	6.53	4.5	6.71	3/9	1.29	0.93
RD (h)	31	8.53	6.5	5.09	5/11	1.33	1.46
V (mm) (all)	72.5	2.3	0.4	5.06	0.1/2	4.97	36.24
V (mm) (roff)	72.5	12.21	9	8.16	7.2/ 15.49	2.21	5.71
RV (mm)	59.3	5.21	2.7	7.11	1.2/ 6.3	3.16	13.92
Iav (mm/h) (all)	13.54	0.6	0.2	0.99	0.1/ 0.7	4.62	33.27
Iav (mm/h) (roff)	10.8	1.63	1.39	0.92	1.05/ 2.77	2.03	4.36
RIav (mm/ h)	6.5	0.62	0.31	0.91	0.12/	3.40	14.33
Im (mm/h) (all)	31.55	1.13	0.3	2.27	0.1/ 1.2	5.28	40.56
Im (mm/h) (roff)	31.55	3.93	3.45	1.96	2.7/ 4.13	2.54	9.22
RIm (mm/ h)	24.98	2.9	1.71	3.89	0.74/ 3.51	3.16	11.72

We can hence conclude that runoff is more likely produced by storms with large volume rather than those with high (average or peak) rainfall intensity. Generally speaking, storms producing runoff last longer, rain greater volumes with greater rainfall intensity.

Focusing on the relationship between storm duration, rained volume and maximum intensity (see scatterplots included in the supplementary documentation), positive linear relationships can be perceived between duration and volume, and between Volume and maximum intensity while, at least for the storms generating runoff, the relationship duration-maximum intensity roughly follows an exponential pattern.

3.1.2. Correlations between storms characteristics and primary atmospheric variables

The relation between the evolution of the primary explanatory variables before the storm starts and the onset of runoff is elusive. Seeking differences between storms resulting or not in runoff (see the boxplots included in the supplementary documentation), we can only highlight that the relative humidity in case of storms producing runoff remains constant before the storm starts, while average, median and third quartile of the other variables tend to slightly decrease in storms producing runoff getting closer to the beginning of the storm. The distributions of the remaining variables for storms producing runoff look slightly more skewed, presenting a larger number of samples at the edge of the distribution. These results, however, are not conclusive, as these subtle statistical variations cannot be associated with certainty to the

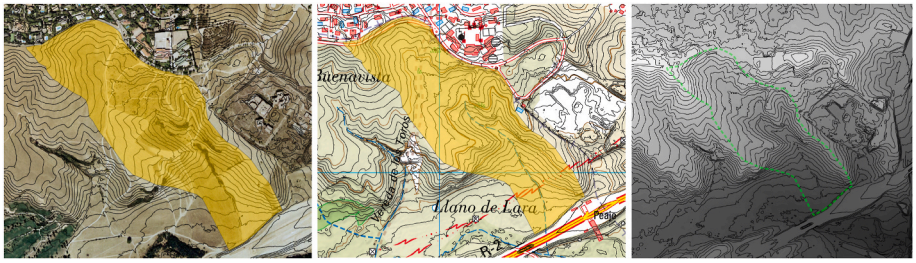

Fig. 3. Case study catchment located in La Solana de Valdebebas.

storms generating runoff (see the boxplots supplementary material).

Yet the linear relationships observed between the storm characteristics and primary atmospheric variables, are not strong, though relative humidity shows positive relations with every analysed storm characteristic, and particularly with the maximum rainfall intensity.

Though no evident non-linear relationships have been revealed between any combination of storm characteristics and primary variable either, some tendencies can be perceived: longer storm duration and large volume (see supplementary documentation, figures SM5 and SM6) are linked with non-extreme pressure and relative humidity and low wind velocity; large volume also looks correlated with temperature lower than average; greater maximum intensity often linked with low wind velocity (see supplementary documentation, figure SM7). Similar conclusions can be drawn when looking at the relationship between runoff related and atmospheric primary variables (see supplementary documentation, figures SM8 an SM.9).

3.2. Data-driven algorithms

3.2.1. Classification task

By applying ML approaches, we can observe that, according to F1-score and ROC metrics (see Fig. 4), LGBM provides better predictions than DNN and that both algorithms perform better when fed on more informative set of inputs (from case a to d). The gap between models, however, reduce as we include more input variables.

As shown in Fig. 4, the model 1a achieves very poor performance (LGBM 0.15 F1 and 0.55 AUC), from which we conclude that this model, only based on a set of non-processed variables observed before the beginning of the storm, is not capable of providing any valuable prediction about the runoff occurrence. Adding the values of predicted volume and maximum precipitation intensity (model 1b) produces a very marginal improvement in performance (increasing the F1 score by 0.05, but leaving the AUC essentially unchanged). With the third approach, where we used precipitation values collected during the first hour of the storm, we instead observe a significant improvement in performance, with an increase of 0.2 for both F1 and AUC scores compared with the previous case. This is a rather interesting result, from which two considerations can be drawn: first, the characteristics of the thunderstorm in an early interval after its onset are informative about the probability of that thunderstorm generating runoff later on; and secondly, these characteristics cannot be properly determined from the variables measured before the onset of the storm, using regression methods.

Finally, with the methodology 1.d, which uses the true precipitation volume and intensities of the storm measured after it is ended, we obtain the best performance (both F1 and AUC exceed 0.8) for LGBM and DNN implementations, as expected considering that these same variables were used to generate the runoff process through the theoretical model that, however, requires a much finer time granularity and additional parameters not available to the ML models. This tells us that the ML approach is potentially able to infer the model parameters from the data, generating good prediction even with less accurate observations, provided that the training data respect the causal relationships predicted by the theory.

Delving into the structure of the predictions, Table 2 presents the true positive and true negative rates.

Except for model 1d, LGBM performs better in predicting the runoff occurrence while, as expected, both perform reasonably well in predicting the negative results.

3.2.2. Prediction task

We now focus on the regression problem, which consists in predicting the runoff volume estimated with Green-Ampt model. Table 3 shows the MSE for runoff regression using both LGBM and DNN predictors for the different approaches previously described.

We see that LGBM provides better performance especially when the real precipitation volume and intensity values are used (model 2d). With the other approaches the performance drops significantly, estimating the mean RV value (see Table 1) with an absolute error of about 2 mm (which yields a MSE of ±4, see Table 3).

LGBM makes it possible to reason on the importance of the features since the relevance of each input can be extracted from the model. Features that result in early stages are given higher importance. We see that the auxiliary variables rainfall volume (V) and maximum intensity (Im), either estimated (as in method 2b) or recorded (methods 2c and 2d) have the highest importance when they are incorporated in the models. Wind velocity measured before the storm has higher performance, followed by pressure and humidity respectively (Fig. 5) while the temperature comes last in terms of importance, except in 2a model (that, however, performs very poorly).

4. Discussion

In this manuscript we aimed at coupling both physically-based and data-driven algorithms. Both approaches have been extensively compared in the literature (Daliakopoulos and Tsanis, 2016; Hsu et al., 1995; Ju et al., 2009; Rauf and Ghumman, 2018; Rezaeianzadeh et al., 2013; Roodsari et al., 2019; Srivastava et al., 2006; Tokar and Markus, 2000; Gharib and Davies, 2021; Kim et al., 2021) and some works show that ML algorithms achieve better performance than physical ones in

Table 2
True positive and negative rates for any model and tested algorithm.

Model	Algorithm	True positive rate	True negative rate
1a	LGBM	0.43	0.59
	DNN	0.00	1.00
1b	LGBM	0.41	0.73
	DNN	0.00	1.00
1c	LGBM	0.58	0.85
	DNN	0.08	0.97
1d	LGBM	0.92	0.98
	DNN	0.92	0.98

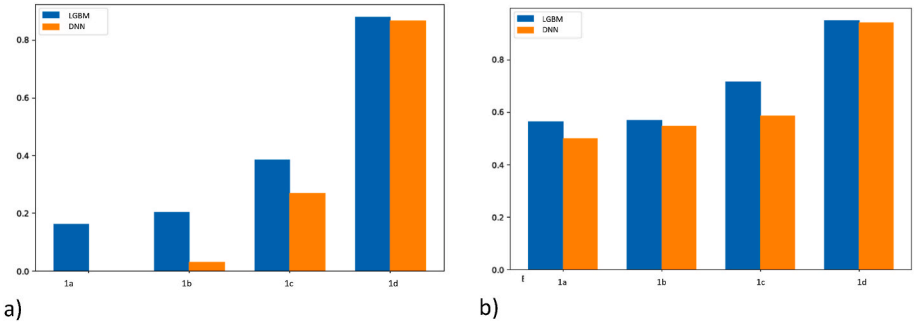

Fig. 4. F1 score (a) and AUC (b) for proposed models.

Table 3
MSE (mm^2) achieved for each model and algorithm.

Model	LGBM	DNN
2a	4.06	4.11
2b	4.09	4.19
2c	4.70	4.71
2d	0.71	1.04

modeling hydrological processes (Dawson and Wilby, 1998; Dibike and Solomatine, 2001; Best et al., 2015; Chaney et al., 2018; Ham et al., 2019; Kim et al., 2020; Shen, 2018; Yang et al., 2017; Zhao et al., 2019). We believe this is not a productive discussion since data-driven models have to learn from physics as the correct pathway for ensuring scientific rigour.

On that idea, we have addressed a simple hydrological problem: predicting the runoff occurrence caused by a storm in a catchment and, if positive, the runoff volume. We used the Green-Ampt infiltration model for estimating the runoff occurrence and, if positive, the runoff volume. We have built data-driven algorithms for predicting the output of the physical modelling (runoff) from the atmospheric variables potentially explaining the rainstorm event before it had started.

We move away from previous experiences in hydrological modelling aided by ML in several planes. First, because we clearly let physics guide the strategy for both input selection and feeding strategy. As many other works, we feed ML algorithms on both stationary (soil-type dependent parameters) and time-dependent (weather records) variables but, in contrast to previous approaches, at two different functional (causal) planes. We force stationary variables to play their role through the physical theory. Following this idea, given the theoretical robustness of the physical law is indisputable, we can thus ensure the data driven algorithm has learned the way the stationary parameters explaining the phenomena are influencing the hydrological process. Engaging the physical theory, we can also be sure the parameters considered are completely and sufficiently explaining the phenomena. In this aspect we differ from previous approaches since many issues can be raised with respect to the validity (causality, completeness and sufficiency) of the parameters commonly used for predicting streamflow in previous works. For example, Kratzert et al. (2019) fed the model on many land, vegetation and soil-dependent parameters together with a set of weather variables, all of them incorporated at the same functional plane. A number of similar cases could also be discussed (see for example Shortridge et al., 2016; Konapala et al., 2020 or Kumar et al., 2021).

Our approach also moves away from previous works with regards to time-dependent input variables. We focus on storm events rather than precipitation time series. Rainfall-runoff processes are determined by storm events and building predictive algorithms on daily (Hao and Bai, 2023; Khosravi et al., 2022; Rezaeianzadeh et al., 2013; Ghorbani et al., 2016; Boucher et al., 2020), monthly (Shabri and Suhartono, 2012; Deo and Şahin, 2016) or even hourly precipitation records (Rozos et al., 2021; Kim et al., 2021) cannot provide the models for predicting the streamflow with the required theoretical consistency. Otherwise, some works feed their models on potentially autocorrelated (precipitation together with atmospheric pressure, temperature or wind velocity, for example) or irrelevant (the effect of potential evapotranspiration to explain streamflow at hourly, or even daily scale can be deeply discussed) variables. Those are the cases of, for example, Gauch et al. (2021), Lima et al. (2016) or Rasouli et al. (2012), among many others. To be sure we avoid the previous issues, we fed the ML algorithm only on a set of primary atmospheric variables potentially triggering rainfall events.

Paying attention to the outputs and the performance of the developed models, we can see that the obtained models fed on the entire set of variables (models 1d and 2d) provided good performance metrics (comparable to the results achieved by the vast majority of previous works presented in the literature), while the algorithms with more interest in predicting in advance (1a, 1b, 1c, 2a, 2b, 2c) did not perform better than the average. Models 1a, 1b, 2a and 2b presented in this work were fed only on time-advanced primary variables, while in models 1c and 2c the recorded precipitation during the first hour was added to the input vector. Though the models 1c and 2c can be applied only 1 h after the beginning of the storm, they are still of interest given that the storms producing runoff last on average 6.5 h, and they would deliver useful predictions once the first hour has passed. The patterns showed by the feature importance can be deemed reasonable when looking at models b, c and d, while they do not look such explicable for models *a*, which suggests the algorithm was not able to efficiently capture the (linear) relationships between the output and the inputs. From the theoretical basis (and also from the observed linear correlation coefficients, though weak), one could expect the relative humidity should have more importance in triggering the storm and, consequently, in explaining the runoff occurrence and volume.

Though we believe the results of some models (particularly, b and c)

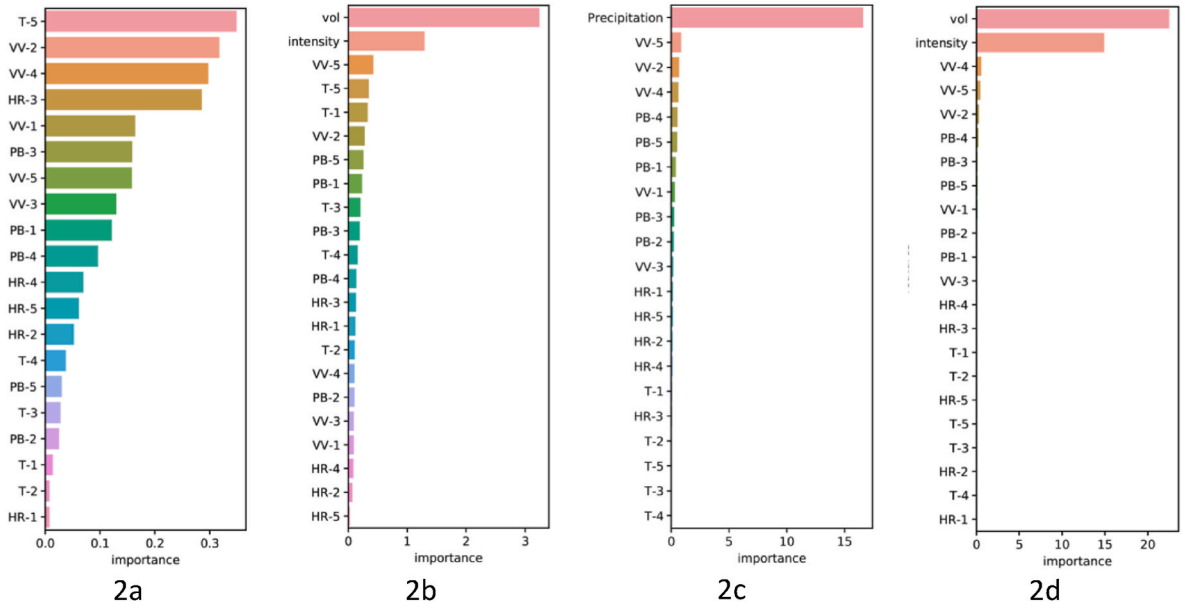

Fig. 5. Features importance in LGBM for each model.

look promising, the idea presented in this work still has field for improvement. We have run simple experiments for setting the required parameters for physical modelling, and better experimental data allowing finer model calibration (on runoff generation and soil physical properties) will undoubtedly drive to more accurate predictions of runoff and better performance of the derived data-driven algorithms. Experience and knowledge ascertain the physical causality between the atmospheric status before the storm starts and the subsequent evolution of rainstorm and, consequently, runoff. Though the data have not robustly supported such statement, we have observed some weak differential patterns between storms producing runoff and not, and some linear correlations between atmospheric variables and storm characteristics that point to those causality. This suggests the dataset is not large enough to let the model capture that pattern. The observed feature importance of models, comprising only atmospheric variables, also suggest the dataset is not large enough to capture the observed correlations between output and inputs. Enlarging the dataset will help overcome those issues.

5. Conclusions

The results achieved in this work have proven the suitability of building effective data-driven algorithms for modelling event-based hydrological phenomena shaped by physical criteria. Letting the data-driven algorithm learn from the output of physical models, and feeding it on primary weather variables, made it possible to interlock both modelling approaches. The algorithm embraces the physical causality among the variables granted by theoretical models and the generalization capabilities of data-driven approaches. The strategy for merging physical and data driven approaches we present in this manuscript represents a novel insight not previously explored by researchers. While previous physics informed ML approaches simply select a set of inputs, often poorly, physically correlated with outputs, we provide a novel insight shaping ML algorithms with physical criteria by letting the ML algorithm learn from physical modelling. The output of the physical model becomes the output for the ML algorithm what represent a clear shift in relation to previous works.

Although the data has not shown clear patterns for the relationship between atmospheric and either storm or runoff-related variables that might have helped the algorithm design, still the developed algorithms have yielded promising performance in predicting runoff occurrence and volume when fed on both primary atmospheric variables before the storm starts and the recorded rainfall volume during the first hour of the storm. Conversely, the algorithms fed only on the set variables recorded before the storms starts have not provided satisfactory performance. Large space for improving has been however detected since the achieved performance is still poor. Checking different algorithms, enlarging the dataset with more data from storms (including more extreme events), considering other variables that may better correlate with storm volume and duration, will likely help obtain better performance.

CRediT authorship contribution statement

Sergio Zubelzu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. **Abdulmomen Ghalkha:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Chaouki Ben Issaid:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Andrea Zanella:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Medhi Bennis:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sergio Zubelzu reports financial support was provided by Spain Ministry of Science and Innovation.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the CHIST-ERA grant CHIST-ERA-19-CES-002. In Spain, this paper belongs to PCI2020-120694-2 project funded by MCIN/AEI/10.13039/501100011033 and the European Union “NextGenerationEU”/PRTR.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2024.120404.

References

Alizadeh, M.J., Nourani, V., Mousavimehr, V., Kavianpour, M.R., 2018. Wavelet-IANN model for predicting flow discharge up to several days and months ahead. J. Hydroinform. 2 (1), 134–148.

Best, M.J., Abramowitz, G., Johnson, H.R., Pitman, A.J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P.A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B.J.J., Nearing, G., Pak, B., Peters-Lidard, C., Santanello, J.A., Stevens, L., Vuichard, N., 2015. The plumbing of land surface models: benchmarking model performance. J. Hydrometeorol. 16, 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1.

Bhasme, P., Vagadiya, J., Bhatia, U., 2022. Enhancing predictive skills in physically-consistent way: physics Informed Machine Learning for hydrological processes. J. Hydrol. 615, 128618.

Blöschl, G., et al., 2017. Twenty-three unsolved problems in hydrology (UPH)–a community perspective. Hydrolog. Sci. J. 64 (1), 1141–1158.

Boucher, M.A., Quilty, J., Adamowski, J., 2020. Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons. Water Resour. Res. 56 (6), e2019WR026226.

Bui, Q.T., Nguyen, Q.H., Nguyen, X.L., Pham, V.D., Nguyen, H.D., Pham, V.M., 2020. Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. J. Hydrol. 581, 124379 https://doi.org/10.1016/j.jhydrol.2019.124379.

Carsel, R.F., Parrish, R.S., 1998. Developing Joint probability distributions of soil water retention characteristics. Water Resour. Res. 24, 755–769.

Chaney, N.W., Huijgevoort, M.H.J.V., Shevliakova, E., Malyshev, S., Milly, P.C.D., Gauthier, P.P.G., Sulman, B.N., 2018. Harnessing big data to rethink land heterogeneity in Earth system models. Hydrol. Earth Syst. Sci. 22 (6), 3311–3330.

Chow, V.T., Maidment, D.R., Mays, L., 1988. Applied Hydrology. McGrawHill.

Daliakopoulos, I.N., Tsanis, I.K., 2016. Comparison of an artificial neural network and a conceptual rainfall–runoff model in the simulation of ephemeral streamflow. Hydrol. Sci. J. 61 (15), 2763–2774.

Darcy, H., 1856. Les fontaines publiques de la ville de Dijon: exposition et application. Victor Dalmont.

Dariane, A.B., Azimi, S., 2018. Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection. J. Hydroinform. 2 (2), 520–532.

Dawson, C.W., Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. Hydrol. Sci. J. 43, 47–66. https://doi.org/10.1080/02626669809492102.

Deo, R.C., Şahin, M., 2016. An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ. Monit. Assess. 188, 1–24.

Dibike, Y., Solomatine, D., 2001. River flow forecasting using artificial neural networks. Phys. Chem. Earth - Part B Hydrol., Oceans Atmos. 26, 1–7. https://doi.org/10.1016/S1464-1909(01)85005-X.

Gharib, A., Davies, E.G., 2021. A workflow to address pitfalls and challenges in applying machine learning models to hydrology. Adv. Water Resour. 152, 103920.

Gauch, M., Mai, J., Lin, J., 2021. The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. Environ. Modell. Softw. 135, 104926.

Ghorbani, M.A., Khatibi, R., Goel, A., FazeliFard, M.H., Azani, A., 2016. Modeling river discharge time series using support vector machine and artificial neural networks. Environ. Earth Sci. 75, 1–13.

Ham, Y.-G., Kim, J.-H., Luo, J.-J., 2019. Deep learning for multi-year ENSO forecasts. Nature 573 (7775), 568–572.

Hao, R., Bai, Z., 2023. Comparative study for daily streamflow simulation with different machine learning methods. Water 15 (6), 1179.
Hsu, K.L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resour. Res. 31 (10), 2517–2530.
Ju, Q., Yu, Z., Hao, Z., Ou, G., Zhao, J., Liu, D., 2009. Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. Neurocomputing 72 (13–15), 2873–2883.
Khosravi, K., Golkarian, A., Tiefenbacher, J.P., 2022. Using optimized deep learning to predict daily streamflow: a comparison to common machine learning algorithms. Water Resour. Manag. 36 (2), 699–716.
Kim, T., Shin, J.-Y., Kim, H., Heo, J.-H., 2020. Ensemble-based neural network modeling for hydrologic forecasts: addressing uncertainty in the model structure and input variable selection. Water Resour. Res. 56 (6), e2019WR026262.
Kim, T., Yang, T., Gao, S., Zhang, L., Ding, Z., Wen, X., Gourley, J.J., Hong, Y., 2021. Can artificial intelligence and data-driven machine learning models match or even replace process-driven hydrologic models for streamflow simulation?: a case study of four watersheds with different hydro-climatic regions across the CONUS. J. Hydrol. 598, 126423.
Konapala, G., Kao, S.C., Painter, S.L., Lu, D., 2020. Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environ. Res. Lett. 15 (10), 104022.
Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Benchmarking a catchment-aware long short-term memory network (LSTM) for large-scale hydrological modeling. Hydrol. Earth Syst. Sci. Discuss. 2019, 1–32.
Kumar, A., Ramsankaran, R.A.A.J., Brocca, L., Muñoz-Arriola, F., 2021. A simple machine learning approach to model real-time streamflow using satellite inputs: Demonstration in a data scarce catchment. J. Hydrol. 595, 126046.
Lima, A.R., Cannon, A.J., Hsieh, W.W., 2016. Forecasting daily streamflow using online sequential extreme learning machines. J. hydrol. 537, 431–443.
Mohammadi, B., 2022. Application of machine learning and remote sensing in hydrology. Adv. Water Resour. 14 (13), 7586.
Mosaffa, H., Sadeghi, M., Mallakpour, I., Jahromi, M.N., Pourghasemi, H.R., 2022. Application of machine learning algorithms in hydrology. In: Computers in Earth and Environmental Sciences. Elsevier.
Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. J. 12, 513.
Neuman, S.P., 1976. Wetting front pressure head in the infiltration model of Green and Ampt. Water Resour. Res. 12 (3), 564–566.
Philip, J.R., 1957. The theory of infiltration: 4. Sorptivity and algebraic infiltration equations. Soil Sci. J. 84, 257–328.
Pourghasemi, H.R., Razavi-Termeh, S.V., Kariminejad, N., Hong, H., Chen, W.F., 2020. An assessment of metaheuristic approaches for flood assessment. J. Hydrol. 582, 124536 https://doi.org/10.1016/j.jhydrol.2019.124536.
Rasouli, K., Hsieh, W.W., Cannon, A.J., 2012. Daily streamflow forecasting by machine learning methods with weather and climate inputs. J. Hydrol. 414, 284–293.
Rauf, A.-U., Ghumman, A.R., 2018. Impact assessment of rainfall-runoff simulations on the flow duration curve of the Upper Indus River—a comparison of data-driven and hydrologic models. Water 10 (7), 876.
Rezaeianzadeh, M., Stein, A., Tabari, H., Abghari, H., Jalalkamali, N., Hosseinipour, E.Z., Singh, V.P., 2013. Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting. Int. J. Environ. Sci. Te. 10, 1181–1192.
Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Physics 1 (5), 318–333.
Roodsari, B.K., Chandler, D.G., Kelleher, C., Kroll, C.N., 2019. A comparison of SAC-SMA and Adaptive Neuro-fuzzy Inference System for real-time flood forecasting in small urban catchments. J. Flood Risk Manage 12, e12492.
Rozos, E., Dimitriadis, P., Bellos, V., 2021. Machine learning in assessing the performance of hydrological models. Hydrology 2022 (9), 5.
Saint-Venant, B., 1871. Theory of unsteady water flow, with application to river floods and to propagation of tides in river channels. French Academy of Science 73, 148–154.
Shabri, A., Suhartono, S., 2012. Streamflow forecasting using least-squares support vector machines. Hydrolog. Sci. J. 57 (7), 1275–1293.
Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resour. Res. 54 (11), 8558–8593.
Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrol. Earth Syst. Sc. 20 (7), 2611–2628.
Srivastava, P., McNair, J.N., Johnson, T.E., 2006. Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed 1. JAWRA J. Am. Water Resour. Assoc. 42 (3), 545–563.
Tikhamarine, Y., Souag-Gamane, D., Ahmed, A.N., Kisi, O., El-Shafie, A., 2020. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. J. Hydrol. 582, 124435, 10.116/j.jhydrol.219.124435.
Thornthwaite, C.W., Holzman, B., 1939. The determination of evaporation fromland and water surfaces. Month. Weather Rev. 67 (1), 4–11.
Tokar, A.S., Markus, M., 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models. J. Hydrol. Eng. 5 (2), 156–161.
Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44, 892.
Wang, Y., Fang, Z., Hong, H., Peng, L., 2020. Flood susceptibility mapping using convolutional neural network frameworks. J. Hydrol. 2020 (582), 124482 https://doi.org/10.1016/j.jhydrol.2019.124482.
Yang, T., Asanjan, A.A., Welles, E., Gao, X., Sorooshian, S., Liu, X., 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour. Res. 53 (4), 2786–2812.
Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., Amagasaki, M., 2022. Capabilities of deep learning models on learning physical relationships: case of rainfall-runoff modeling with LSTM. Sci. Total Environ. 802, 149876.
Zanella, A., Zubelzu, S., Bennis, M., 2023. Sensor networks, data processing, and inference: the hydrology challenge. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3318739.
Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: a review. J. Hydrol. 589, 126266.
Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, Changjie, Xi, L., Q, Guo Y., 2019. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 46 (24), 14496–14507.
Zhong, L., Lei, H., Gao, B., 2023. Developing a physics-informed deep learning model to simulate runoff response to climate change in alpine catchments. Water Resour. Res. 59 (6), e2022WR034118.
Zuo, G., Luo, J., Wang, N., Lian, Y., He, X., 2020. Two-stage variational mode decomposition and support vector regression for streamflow forecasting. Hydrol. Earth Syst. S. 24 (11), 5491–5518.