IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 August 2023, accepted 20 September 2023, date of publication 25 September 2023, date of current version 5 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3318739

TOPICAL REVIEW

Sensor Networks, Data Processing, and Inference: The Hydrology Challenge

ANDREA ZANELLA[1,2], (Senior Member, IEEE), SERGIO ZUBELZU[3], AND MEHDI BENNIS[4], (Fellow, IEEE)

[1]Department of Information Engineering, University of Padova, 35122 Padua, Italy
[2]Consorzio Nazionale Italiano per le Telecomunicazioni (CNIT), 43124 Parma, Italy
[3]Departamento de Ingenierí-a Agroforestal, Escuela Técnica Superior de Ingenierí-a agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain
[4]Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland

Corresponding author: Andrea Zanella (andrea.zanella@unipd.it)

This work was supported by the CHIST-ERA Grant CHIST-ERA-19-CES-002.

ABSTRACT In the last years, many European countries have experienced the effects of climate change, in the form of a scarcity of drinking water resources, prolonged periods of drought, and extremely heavy rainfall, with unprecedented dramatic environmental, economic, and social costs. Therefore, understanding, modeling, and predicting the movement and distribution of water on Earth, and effectively managing water resources are problems of paramount importance. In this article, we discuss the fundamental role that sensing technologies, data processing algorithms, and inference based on machine learning techniques can have in modern hydrology and the many challenges that still need to be addressed to improve the accuracy and reduce the complexity of current hydrology models. More specifically, we overview the main solutions proposed in the literature to monitor, analyze and predict hydrological processes, and present a selection of results obtained from empirical data sets to ground the main concepts and substantiate the dissertation. Finally, we conclude our article by discussing open problems and possible avenues for future research.

INDEX TERMS Hydrology, sensor, machine learning, ICT, rainfall forecast, communications technology -> wireless sensor networks -> event detection, geoscience and remote monitoring -> remote sensing -> remote monitoring, instrumentation and measurement -> monitoring -> water monitoring, computational and artificial intelligence -> artificial intelligence -> prediction methods -> predictive models.

I. INTRODUCTION

Hydrological systems are complex networks of interlinked and highly variable processes, including precipitations, water infiltration and evapotranspiration, and surface and groundwater movement. Over the years, these processes have been studied, yielding a plethora of robust physical theories and models to represent and predict each process. For example, physical theories based on non-linear partial differential equations can be easily found for precipitation, evapotranspiration [1], soil water flux [2], [3], [4], [5] or open-channel flow [6]. Besides such theoretical models based on physical properties, a number of widely used empirical models have also been developed (e.g., [7], [8], [9]).

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Fadda.

These somewhat classic models have recently been enriched by data-driven techniques. The machine-leaning approach has become possible due to the increased availability of datasets, which, in turn, is due to developments in measurement and monitoring technologies, including satellites and unmanned aerial vehicles for geographic imagery, radar, and advanced *in-situ* measurement devices to collect not only meteorological data but also parameters related to terrain composition and landscape shape. Taking advantage of these technologies, hydrologists have made great efforts to improve their understanding of physical processes and refine theories using accurate experimentalmeasurements.

Despite the seemingly mature theories and advanced tools available for data collection and processing, and for solving complex non-linear systems of partial differential equations, large-scale theoretical models are still prohibitively complex to handle with classical approaches. In addition, hydrological models are often hyper-parameterized to account for the many different factors that affect hydrological processes. The accuracy of such models, therefore, depends heavily on the ability to correctly characterize their parameters, which is a daunting ask considering that these parameters can be very sensitive to geographic and climatic conditions. In contrast, empirical approaches seem easier to use, but they lack generality and, despite their wide use in practical applications due to their (relative) simplicity, they are not currently considered as definitive solutions to the problem of modeling hydrological processes. Therefore, a complete understanding of the interactions among the physical processes underlying the hydrological phenomena remains elusive [10].

Hydrologists are well aware of these difficulties. Recently, a set of 230 authors, including well-known hydrologists and scientists from other disciplines related to hydrology, highlighted in [11] some relevant unsolved issues, including the inability of existing hydrological laws to adequately model processes at different scales, the need for innovative technologies for data collection and modeling, the implications of using historical or synthetic data, the need to reduce the uncertainty in the models' structure, parameters, and input in order to improve hydrological predictions. In fact, most hydrologists believe that high-quality experimental data are still lacking and suggest that a better understanding of hydrology should be based on high-quality experimental data rather than better modeling approaches [12].

Therefore, despite the recent advances in hydrology and related disciplines, scientists have not yet reached a clear understanding of the physical theories, nor have they found sufficiently general empirical models.

To help improve the understanding of hydrological processes and make a clear step forward, hydrologists emphasize the importance of interdisciplinary approaches, calling for greater integration within the subfields of hydrology and with other water-related disciplines [12].

In this article, we argue that information and communication technologies (ICTs), together with machine learning (ML) methodologies, are instrumental in developing more accurate hydrological models and effective water management strategies and interventions. In particular, the following three crucial aspects can be vastly improved through the proper use of ICT and ML.

1) *Data collection*. Today, data are generated by weather stations or soil-moisture and water-level sensors scattered throughout the territory, and data collection is often too sparse in time and space to provide a good picture of the correlations between the measured processes. In addition, data are mostly produced at a constant frequency, regardless of weather conditions, meaning that most of the data have low (or no) information content for modeling (critical) hydrological phenomena. Proper planning of the measurement stations in the territory, the use of satellite images, and the development of adaptive sampling techniques can dramatically improve the quality of collected data while reducing their size. In parallel, numerical methods can be used to generate synthetic data sets that, in turn, may be useful to pre-train machine-learning models, thus reducing the need for empirical measurements and, hence, for dense sensing networks.
2) *Data processing.* Once collected (or generated synthetically), data need to be processed to refine hydrological models and develop accurate predictions of extreme events. As mentioned above, however, the available data are often unbalanced, noisy, and uneven, that is, of low quality. Therefore, refinement techniques are needed to reduce the size of data sets, remove outliers, align time series generated from different sources, and better balance the data to have more representative sets before feeding them to ML algorithms. Although these operations are actually common in data-driven approaches, the specificity of the hydrological domain requires particular care and dedicated solutions, which have yet to be studied in depth.
3) *Data visualization.* Finally, once information has been extracted from the available data, it is essential to develop methodologies to facilitate the reading and interpretation of patterns and related inferences. Hyetographs, probability distributions, box plots, and other common data visualization tools used in the field are accessible only to experts with a technical background, who have the skills to correctly read and interpret the models' outcomes. However, decision-making in the water management domain involves a number of stakeholders, not all equipped with the background to correctly interpret the models, or their outcomes, using such technical representations. Therefore, it is necessary to develop new solutions based on modern data visualization techniques such as, e.g., augmented reality, 3D visual models, and dynamic maps.

This article is intended as an introductory guide to the state of the art and to the many open challenges that still need to be addressed to improve current hydrology models. In particular, we focus on two specific questions:

> *Is it possible to determine the level of water along the main discharge river of a catchment based on the data collected before, during, and after a storm? And is it possible to predict the water level well in advance?*

These abilities are indeed critical to controlling water locks before, during, and immediately after a storm to manage tributaries and the main channel so as to avoid or, at least, limit the risk of flooding. Note that, the problem is much more complex than just precipitation estimation, which is only one of the elements that determine the water level in the basins. Currently, weather radars provide models with high-density spatial information on precipitation. However, they involve high investment and operational costs and would provide only a partial answer to estimating the variability of hydrological processes, being unable to capture other aspects that impact the time-space variability of the water levels, such as water infiltration rate, runoff (i.e., surface water) percentage, and flood routing over the ground.

In table Tab. 1 we provide a research comparison table with the references used in this manuscript.

In summary, this paper provides the following contributions:

(i) a high-level introduction to the fundamentals of hydrology to equip ICT practitioners with the basic knowledge needed to understand (and appreciate) the challenges offered by the hydrological domain;
(ii) a review of the leading solutions proposed in the literature regarding the problem of hydrological data detection and processing;
(iii) a selection of results obtained in real-world scenarios to ground the discussion and exemplify possible solutions;
(iv) a reasoned debate on open challenges and possible approaches that can be adopted to tackle them.

This manuscript is mainly intended for scientists from the computing and networking domains, with the aim of raising awareness of hydrological issues seeking the aforementioned interdisciplinary insight. With this ambition, after a quick introduction to the basics of event-based hydrological processes (Sec. II), we overview the state of the art in monitoring and modeling event-based hydrology in Sec. III and discuss the open challenges in Sec. IV. Finally, Sec. V concludes the paper with a set of bullet points and lines of action for sensing, modeling, and data processing to help improve the understanding of hydrological processes.

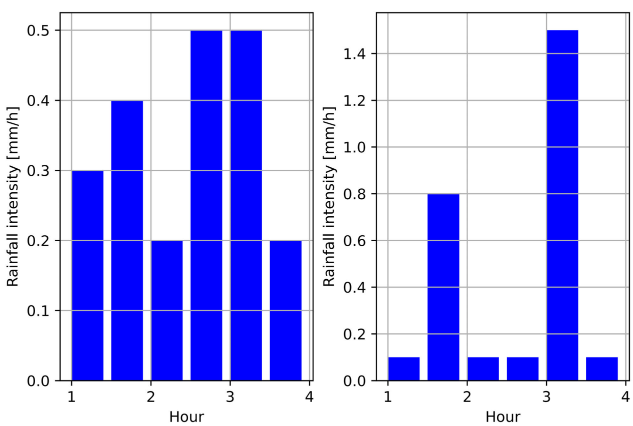


FIGURE 1. Example of hyetographs for two different storms in Madrid catchment, with similar aggregate precipitation volume.

II. BASICS ON EVENT-BASED HYDROLOGICAL PROCESSES

The reference scenario considered in this work is a hydrological catchment, i.e., a portion of land where water inputs (basically, rainfalls) feed water outputs (rivers, evaporation, and evapotranspiration processes) and change the internal variables (ground humidity, aquifers level). Given the physical contour of the catchment, the transfer function within it is simply given by

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = I(t) - O(t) \,, \tag{1}$$

where $S(t)$ is the system's internal state (i.e., the total amount of under-surface water), while $I(t)$ and $O(t)$ are the aggregate rate at which water enters and exits the system, respectively.

When focusing on event-based hydrology, it is customary to identify the inputs with the intensity of precipitations and the outputs with the water flow rate through the downstream end-point of the catchment's discharging channel. The period of interest goes from the beginning of a storm (i.e., the first instant t_0 for which $I(t_0) > 0$) and the time when the surface runoff has completely gone out through the downstream endpoint of the catchment. (i.e., the minimum time $t_{end} > t_0$ at which the output flow rate $O(t)$ stabilizes at the value prior to the storm). Notice that, evaporation, evapotranspiration, and aquifers' recharge processes can be neglected at this time scale. Therefore, the variation of the system's internal state is basically due to rainfall infiltration processes. For $t_0 \le t \le t_{end}$, the variation of the internal state, i.e., $S'(t) = dS(t)/\mathrm{d}t$, is equal to the infiltration rate, and $O(t)$ is the difference between rainfall and infiltration rates delayed by the transient time τ, i.e., the time the flow takes to reach the downstream-end point: $O(t) = I(t-\tau) - S'(t-\tau)$.

A. STORM EVENTS

Storms trigger water input processes and represent our first spatial and temporal variability source. Storm events often appear and develop unevenly within a catchment. They can start at different instants, and have variable duration and intensity over time, thus delivering different precipitation volumes in each area.

A typical way to represent a storm event is by means of hyetographs, which are histograms of the precipitation intensity over discrete time intervals. The rainfall intensity is the amount of rain falling during a reference period of time (e.g., 1 hour). The amount of rainfall is measured with reference to the height that would be reached by rainwater inside a cylindrical vessel with a unit base area placed on the ground (rainfall depth) and is conventionally expressed in mm. The rainfall intensity is hence expressed in terms of depth units per unit of time, i.e., [mm/h].

Examples of the hyetographs obtained from Madrid's weather stations on two different days are shown in Fig. 1. Notice that the aggregate volume of precipitation during the whole day is roughly the same for the two storms, while the hourly intensities are very different.

In Fig. 2 we report other examples of hyetographs obtained in different conditions: the plot on the left-hand side reports the recorded hyetographs of storm events collected by a weather station located in Madrid. The middle plot reports the hyetographs of only the storm events generating runoff

TABLE 1. Summary of analyzed manuscripts sorted by year and grouped by themes.

Year	Use of remote sensing (satellite images, UAVs) for measuring hydrological processes	Review and analysis of modelling and monitoring techniques for hydrology	Use of remote sensing and ML techniques for modelling hydrological processes.	On-site monitoring of hydrological processes	Hydrological processes monitoring aided by ICT tools	Spatiotemporal variability of hydrological processes	Machine learning for hydrological modelling	Optimal monitoring theories	ML techniques for image processing
1856		[2]							
1911		[5]							
1931		[4]							
1932		[7]							
1933		[8],[9]							
1939		[1]							
1957		[3]							
1976		[14]							
1977								[108], [109]	
1979				[77]					
1980		[13]							
1987								[110]	
1988		[15]							
1997									[46]
1998	[24]								
2001								[119]	
2002								[111]	
2003				[89], [90]				[113], [125]	
2004		[105]		[91]	[66]			[112]	
2005							[93]	[117], [123]	
2006					[64], [68]		[135]	[124]	
2007	[31]	[107]		[58]	[65]				
2008	[30]			[49], [55]				[128]	
2009				[57]				[122]	
2010				[54]					
2011	[23]			[59], [60]					
2012		[145], [140]		[48]		[71], [152], [153], [154]	[96]	[116]	
2013				[52], [61], [78]					
2014	[16]			[51], [53]		[70]			
2015	[27]	[37]		[50], [76]		[72]			[45]
2016	[25], [26], [29], [35]					[73]		[114], [118]	
2017	[17], [18]	[11], [12]		[56]			[151]	[115]	
2018		[10], [137]					[87], [88], [101], [103], [150]		
2019	[19], [20], [28], [32], [34], [79], [81]			[62]		[74], [155]	[102], [130], [133]	[126]	
2020	[36],[106]	[142]				[80], [147]	[86], [92], [94], [95], [97], [98], [99], [132], [139], [138]	[120], [121]	
2021	[21], [22],[33], [129]	[41], [141], [144]		[63]	[67], [69]		[85], [144], [149]		[47]
2022	[38]		[39], [40]				[82], [83], [84], [100], [104], [132], [136], [131], [143], [149]		[42], [43], [44]
2023					[75]		[130]		

(which are a subset of the previous ones). Finally, the right-hand side plot reports the runoff hyetographs obtained by using theoretical models (namely, Green-Ampt model, see later) during storm events generating runoff. Comparing the first and second plots, we can observe that the hyetographs for storm events that result in runoff are typically skewed

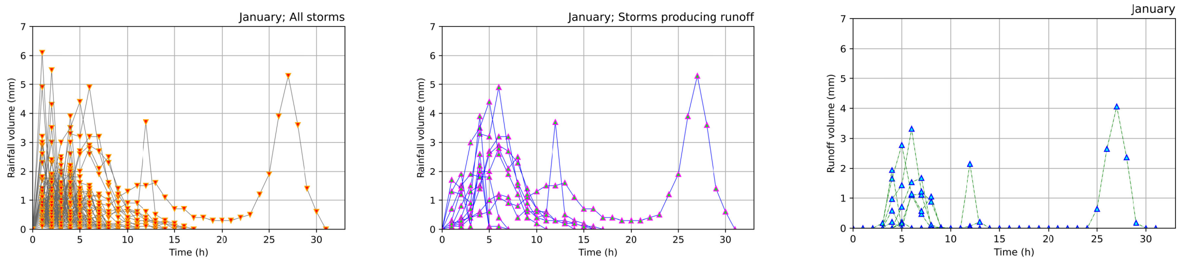


FIGURE 2. **Example of hyetographs for synthetic (left-hand side and middle figures) and empirical (right-hand side figure) precipitation intensities. Middle and right-hand side figures correspond to storm events that generate run off.**

to the right, denoting a tendency for the storm to increase in intensity after some time from the start of rainfall. The graph on the right is similar, but much less defined because the infiltrated rainfall has been removed.

In many practical cases, weather stations provide only the aggregate value of precipitation over long time intervals (hours or even a whole day). In this case, fine-grained hyetographs can be estimated from the maximum rainfall intensity (or from the aggregated precipitation volume) and the storm duration, assuming a normal-type distribution of the precipitation intensity over time. However, often the storm duration is known only approximately. In addition, as can be observed from the hyetographs of Fig. 1 and Fig. 2, the hyetograph distributions can be quite far from normal, particularly for thunderstorms that produce runoff. Therefore, the hyetographs obtained with this technique are often quite inaccurate.

The main issue with storm prediction is the spatial variability within the catchment. As previously noted, precipitation is highly space-dependent and evolves differently depending on site-specific factors. This is apparent from Fig. 3, which shows the cumulative probability distribution (CDF) of the maximum thunderstorm intensity (left-hand side) and thunderstorm volume (right-hand side) obtained from data collected in different months of the year from six meteorological stations, located in Madrid, up to 20 km apart. We can see that the CDFs for the different stations in each graph do not perfectly overlap, denoting a difference both in the maximum rain intensity and aggregate precipitation volumes among the measurement locations.

B. INFILTRATION/RUNOFF AND WATER ABSTRACTIONS

The soil infiltration capacity across the catchment and along the storm duration mainly determines the water discharging rate at the downstream endpoint. At the catchment's scale, it mainly depends on the land use, which may vary from completely impermeable areas (urban paved areas) to permeable uses (e.g., open green spaces or meadows).

The specific infiltration at each (permeable) soil type can be predicted with different models. Some notable models are those proposed by Philip [3], based on Richards' equation [4], and by Green-Ampt [5]. Denoting by $f(t)$ the infiltration rate at time t, and by $F(t)$ its integral (i.e., the aggregate volume of infiltrated water up to time t), Philip's model gives:

$$f(t) = K_s + \frac{S_o}{2\sqrt{t}}; \tag{2}$$

while the Green-Ampt model gives:

$$f(t) = K_s \left(\frac{\varphi \Delta\theta}{F(t)} + 1 \right), \tag{3}$$

$$F(t) = K_s t + \varphi \Delta\theta Ln \left(\frac{F(t)}{\varphi \Delta\theta} + 1 \right). \tag{4}$$

The above equations are all parameterized: K_s and $\Delta\theta$ are soil-dependent parameters (including the initial soil water content at $t = t_0$), while S_o and φ are soil-type dependent non-linear functions. The most widely used forms for such functions are those based on Van Genutchen, [13] and Mualem's models [14], which depend on seven soil-type specific additional parameters (initial, saturated and residual soil water content, the so-called soil *tortuosity* and three additional fitting parameters). This means that, to determine the infiltration rate at the catchment's scale with the most common physically-based models, we need to tune at least seven parameters, which are highly dependent on the soil's physical properties. Though there are some databases providing average values for those parameters for a set of soil categories, soils are complex and highly variable in space, and small differences in parameters' estimates can yield large errors due to the high sensitivity of the models to such parameters. As an example, Tab. 2 shows the parameters' estimates (θ_s and θ_r stand for the saturated and residual soil moisture, two required parameters for solving the aforementioned infiltration functions) based on both field measurements and laboratory tests, for a sample of 64 points within a small land with homogeneous use in the Madrid's catchment (see Fig. 4), compared to the average values retrieved from the most widely used database for those parameters [15].

The data show that: 1) tabular parameter values differ significantly from the experimental ones and 2) experimental values show a large variability even if samples were collected in a small catchment with homogeneous land use.

In conclusion, the estimation of infiltration rate during thunderstorms is influenced by the spatial variability of rainfall intensity and land use. Infiltration increases spatial and

TABLE 2. Comparison of mean, coefficient of variation (CV), maximum and minimum values of the infiltration model's parameters obtained in the lab (from the sampling points at the catchment presented in Fig. 4) and taken from the literature. The error is the normalized difference of the empirical mean values with respect to those in the literature.

Metric	Source	Mean	CV	Max	Min
θ_s (cm^3/cm^3)	Carsel and Parrish (1988)	0.414	0.02	0.41	0.41
	Laboratory	0.334	0.255	0.572	0.097
	Error	19%	-1175%	-40%	-76%
θ_r (cm^3/cm^3)	Carsel and Parrish (1988)	0.05	0.131	0.065	0.045
	Laboratory	0.023	1.034	0.132	0.00
	Error	54%	-689%	-103%	100%
k_s (cm/min)	Carsel and Parrish (1988)	1.77	0.47	2.43	0.74
	Laboratory	2.045	0.767	8.08	0.363
	Error	-16%	62%	-231%	51%

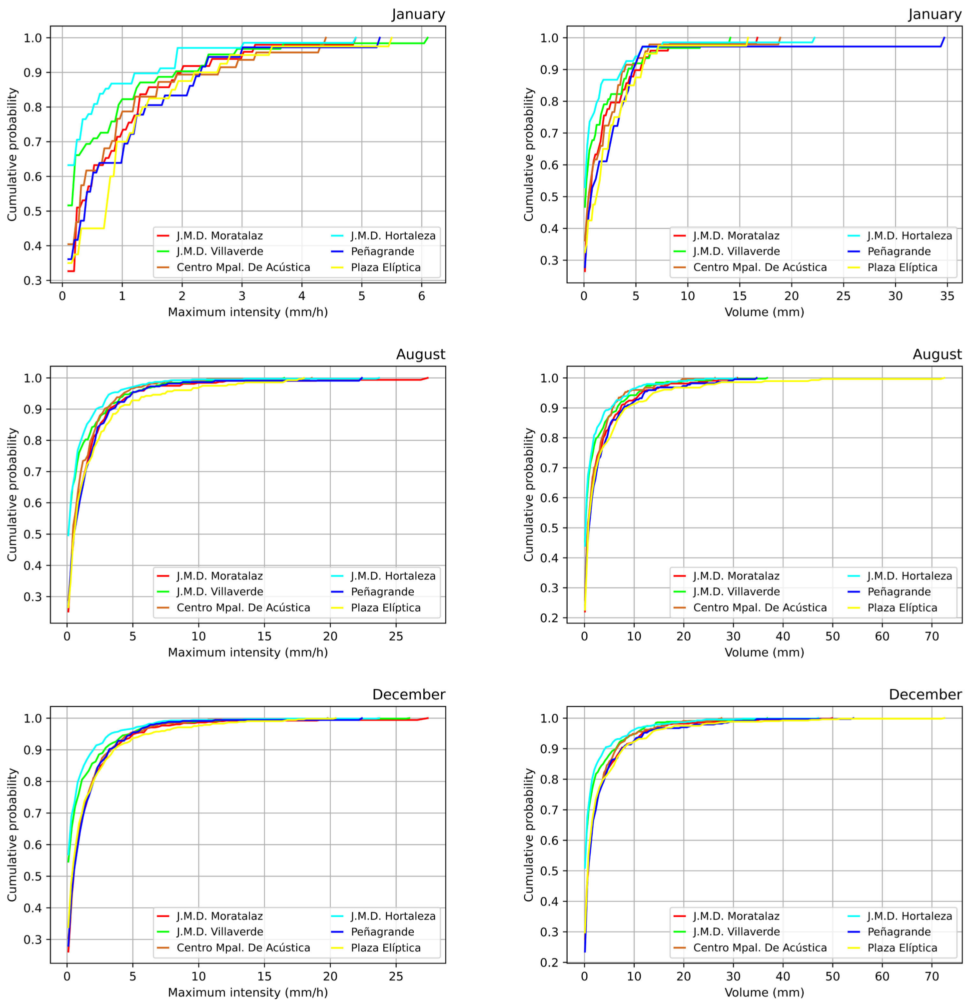


FIGURE 3. CDFs of the peak rain intensity (left-hand side graphs) and aggregate precipitation volume (right-hand side graphs) for six weather stations in Madrid's catchment, in different periods of the year.

temporal variability of the process to be monitored/predicted (water level in the river over time). Moreover, the theoretical models are complex and highly sensitive to a number of soil-dependent parameters, which are also difficult to estimate.

C. FLOOD TRANSIENT

The part of rainfall that does not infiltrate into the ground flows downstream towards the channel courses. In the steady-state, when intermediate water storages are full, this surface precipitation provides the so-called *runoff contribution* to discharging channels. It starts at the same time as the storm, or later (depending on the location of the storm, its intensity, and the ground moisture), and usually continues for a period after the end of the storm, because of the time it takes for runoff to reach the channel courses.

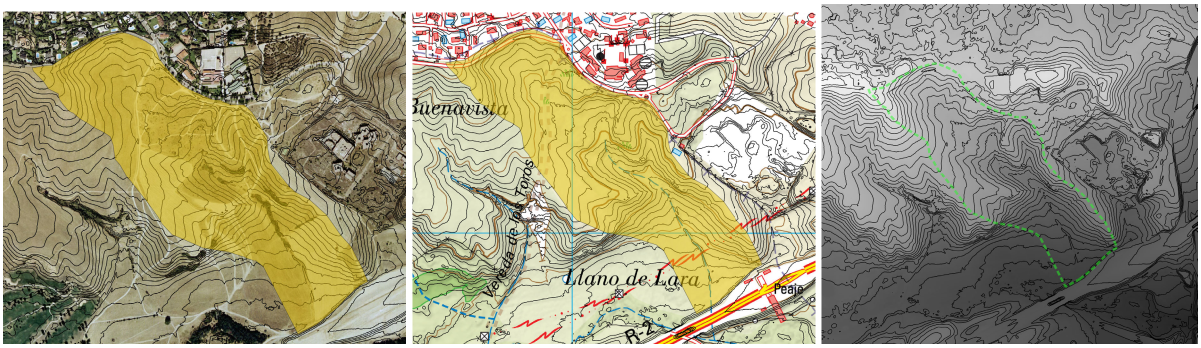


FIGURE 4. Small (300.000 m^2) catchment in Madrid.

This phenomenon is well described by means of hydrographs, i.e., graphs showing the rate of water flow past specific points in a river (discharge) versus time. Input hydrographs are heterogeneously distributed across the main course: there can be discrete discharging points where tributary courses reach the main channel and continuous or discrete contributions to tributary rivers.

The literature has traditionally addressed such problems with either conceptual models, as that presented in (1), or physical theories which often yield non-linear systems of partial differential equations without general analytical solutions. As an example, we report the 1D Saint-Venant equations [6], which describe the velocity v along the spatial coordinate x of a shallow-water discharged flow (Q) over a cross-section of area ω with depth y, driven by the available energy head $I_o - I$:

$$0 = \frac{\partial Q}{\partial x} + \frac{\partial \omega}{\partial t}; \tag{5}$$

$$\frac{1}{g}\frac{\partial v}{\partial t} + \frac{v}{g}\frac{\partial v}{\partial x} + \frac{\partial y}{\partial x} = I_0 - I. \tag{6}$$

Such approaches either require rough simplifications (for example assuming uniform flow) or imply complex mathematical formulations and great computation efforts. In any case, a number of simplifications are required to estimate the transient of the hydrograph across the catchment since it is highly affected by many physical (territorial) factors not managed by these models, particularly in large and complex catchments. Moreover, solving physical theories as those expressed by Saint-Venant's equations faces a number of computational issues that, together with the complexity of parametrization at the catchment scale, or the inability to deal with stochastic processes, hinder the use of physically-based models in practical applications. On the other hand, there is also a great variety of empirical models that, however, lack generality, so their solutions are often very imprecise.

III. STATE OF THE ART IN MONITORING AND MODELING EVENT-BASED HYDROLOGY

In this section, we focus on event-based hydrology and, more specifically, on the fundamental phenomena that are typically considered in this context, namely precipitation, infiltration, and flood routing. The problem of measuring, monitoring, and modeling these processes has been addressed by a number of studies. In the following, we outline the most relevant approaches presented in the literature to provide a broad and general view of the subject.

A. SENSING TECHNOLOGIES, MEASUREMENT AND MONITORING

In recent years, hydrology has greatly benefited from developments in the fields of wireless communications, unmanned vehicles, satellite imagery, and computational techniques. In the literature, however, these technologies have been used for the analysis of individual processes involved in event-based hydrology, rather than the system as a whole. A detailed study with an interesting discussion can be found in [10], which presents the main conclusions from the Measurements and Observations in the XXI century Working Group of the International Association of Hydrological Sciences (MOXXI) involving the majority of the most relevant hydrologists around the world.

Satellite images have provided accurate measurement of many variables of great interest for hydrology as soil water content [16], [17], [18], [19], [20], [21], [22], evapotranspiration [23], [24], [25], streamflow [26], [27], [28], [29], [30], [31] or precipitation [32], [33], [34], [35], [36], [37]. In fact, satellites can deliver images with high spatial and temporal resolution. From them, precious information about hydrological processes can be extracted by image processing techniques, often based on ML. A comprehensive review of such techniques with application to remote sensing can be found in [38], and some examples are given in [39] and [40]. In addition to satellites, unmanned aerial vehicles can also provide images from a lower altitude than the satellite, thus with higher spatial and temporal resolution. Image processing techniques have been developed also for these images as discussed, e.g., in [41].

Although these techniques have contributed to advance remote sensing for hydrological modeling, there is still a large space for improvement. Deep learning shows great promise in this direction [42], [43], [44] and can potentially provide hydrologists with powerful tools for patterns recognition in hydrological processes [45], and image understanding [46], among others. The suitability of computer vision algorithms for hydrological modeling has been recently discussed in [47], but the algorithms have not yet been extensively tested on hydrological processes.

While the larger availability and accuracy of aerial sensing has certainly paved the way to new measurements and data processing techniques, on-ground measurement technologies have also experienced significant advances. These advancements are also due to the improved performance and efficiency of communication technologies. In fact, the possibility of transmitting data over long distances with little energy expenditure, or the use of optical fibers as distributed sensors, have allowed hydrologists to collect data from remote areas, e.g., to accurately determine the soil water content [48], [49], [50], [51], [52], [53], [54], the streamflow [55], [56], [57], [58], [59], [60], or the precipitation [61].

Some such devices make it possible to overcome the limits of point measurement devices, such as the classical rain gauges, and collect information on wider areas around the actual location of the devices. For example, cosmic-ray neutron sensing [62] makes it possible to estimate the soil water content in large volumes of soil, while radar systems [63] allow for precipitation measurement over wide areas.

Technologies designed and developed for other purposes may also prove useful in hydrology. For example, [64], [65] investigated how the wireless cellular systems used for communication can be exploited as a widely distributed sensing network for atmospheric phenomena by correlating the received signal level with surface rainfall.

Though on-ground isolated devices can provide great accuracy in measuring some processes, they can hardly deal with hydrological spatial variability, so the ability to install and manage monitor networks and develop effective data processing techniques has become of paramount importance. The development of wireless sensor technologies has allowed hydrologists to deploy dense monitoring networks [66], [67], [68], [69]. In particular, wireless sensors networks have been used to monitor different hydrological processes in isolation as for example spatiotemporal variability in soil moisture [70], [71], catchments [72], or preferential flows in forested catchments [73]. Some applications for wireless hydrological modeling fall within the so-called Internet of Underground Things [74].

Despite the possibilities offered by the existing wireless technologies, some issues related to maintenance (mainly, energy-related issues) still limit the possibility of deploying dense-enough networks to reach the desired spatial accuracy, as discussed in [75].

Weather radars have also emerged to provide hydrologists with highly valuable information on precipitation. Radars can be used for retrieving highly detailed spatial information (for example portable X-band weather radars). Many studies have addressed the use of precipitation radar information for hydrological modeling and prediction. Some comprehensive studies can be found in [76] or [77], while an interesting discussion on the real feasibility of radars in practical applications is presented in [78].

Finally, is worth mentioning that different types of sensors can be combined in a synergistic way. For instance, the combination of ground sensors and satellite data has driven better modeling tools [29], [79], [80]. However, merging information from different sources can also raise data-related issues, as discussed, e.g., in [81].

B. DATA-DRIVEN APPROACHES

In recent times, hydrological modeling has benefited from the enhanced performance of data-driven models and the wider availability of data to train such models. Data-driven models open up interesting opportunities to deal with most unsolved modeling issues related to hydrology. Scientists from different disciplines have applied data-driven algorithms to the different processes involved in hydrology. In addition, some initial attempts have been made to develop simplified models for the entire hydrological system.

The amount of papers on this topic is enormous and constantly increasing. Some recent surveys are [82], [83], [84], and [85], but the research in this area is flourishing and the state of the art is evolving rapidly. In this section, we attempt an introductory overview of the different problems that have been addressed with data-driven approaches.

1) RAINFALL PREDICTION

As previously mentioned, the reliability of weather records used to feed hydrological models is an open issue and data-driven algorithms have been used to improve the quality of precipitation records. A certain effort has been dedicated to modeling precipitation events in general, while the prediction of specific storm events (e.g., those generating runoff) has not received as much attention.

Many works have applied a wide range of data-driven algorithms combined or not with different data pre-processing techniques for rainfall forecasting. For example, [86] developed a complex ensemble model hybridized with both random forest and kernel ridge regression for monthly rainfall forecasting. Their approach required first to factorize rainfall time series into their respective intrinsic mode functions using an empirical model decomposition. Once the significant lags of each intrinsic mode function and the residual were identified, forecasting techniques aided by random forest were defined for both and, finally, the kernel ridge regression model was adopted, combining the forecast intrinsic mode functions and the residuals to generate the rainfall estimates.

Following a similar stage-based strategy, the work [87] proposed a mixed method combining the *Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors* (SARIMAX) method with the least squares support vector machine (SVM). They first decomposed the rainfall time series by means of a wavelet transformation, then used a

SARIMAX model to fit the linear components of the rainfall time series, and applied a least square SVM (LSSVM) to model the residuals, supposedly containing the nonlinear relationships. Finally, they used both predictions (sub-series from SARIMAX and residuals from LSSVM) to rebuild the original rainfall time series.

Others sought to go further in short-term prediction, developing algorithms for sub-daily precipitation estimation from daily data from a dense monitoring network [88]. Seeking to overcome the discrepancy between (daily) precipitation and (hourly) streamflow data, they split up the first to hourly scale using a multivariate approach to feed a physical model. The results however were not impressive and they highlighted the importance of dense monitoring networks for improving the accuracy of the proposed multivariate method for daily rainfall disaggregation.

The short-term prediction of rainfall evolution is of paramount importance in event-based hydrology and, though accurate techniques based on now-casting are currently available [89], [90], [91], [92], they need real-time dense spatial information gathered from radars, which is not always available in locations visited sporadically by satellites. Radars can provide high-quality data with detailed spatiotemporal information, but the now-casting techniques used for data processing are not exempt from methodological issues. To quantify the uncertainty of predictions, [93] used a Bayesian joint probability nowcasting scheme providing not only forecasts but also a measurement of their uncertainty.

2) RUNOFF PREDICTION

As explained in Sec. II-B, accurately modeling rainfall-runoff processes is a problem of particular complexity because of the several different processes that are involved and their spatial heterogeneity. This topic has attracted great attention, and a number of data-driven algorithms have been proposed in the literature, either focusing on particular processes or on the whole system. These approaches can be grouped into several classes, as explained next.

a: PURE AUTOREGRESSIVE MODELS FOR FORECASTING STREAMFLOWS AT DIFFERENT TEMPORAL SCALES

In [94], the authors used artificial neural networks (ANNs), support vector regression (SVR), and Multiple Linear Regression (MLR) methods coupled with the Grey Wolf Optimizer algorithm (GWO), with the intent of predicting monthly streamflow in the Nile river. The input vector consisted of K consecutive samples of the different observable variables, where K was selected so as to maximize the autocorrelation functions of the variables, in order to capture the time dependencies. They found that the best performance of the model was obtained with streamflow records lagged apart by 3 samples. In [95], instead, they used a two-stage decomposition combining variational mode decomposition and SVR to efficiently predict monthly streamflow. In [96] the authors also defined an autoregressive approach checking different algorithms (least square SVM, SVM, ANN, and ARIMA) for predicting monthly streamflow in two Chinese catchments, defining models from 1 to 11-month lags. They found the models fed with either 7 or 8 monthly lagged records showed better performance.

b: DATA-DRIVEN MODELS BUILT UPON APPARENT (PHYSICAL) CAUSALITY CRITERIA

Several works depart from autoregressive models and develop data-driven models shaped by physical relationships, given there is physical causality between the dependent variable (y) and a set of independent variables (x_j) as expressed in (7):

$$y_t = f\left(y_{t-i}, x^j_{t-i}\right), \quad 0 \leq i < t \tag{7}$$

For example, [97] and [98] used neural networks for predicting flood susceptibility areas using a set of different input variables (topography, vegetation-related, rainfall, land use, slope, lithology, etc.) that are undoubtedly related with the outputs (though the authors do not explain the physical causality with flood susceptibility). As expected, they achieve good performance in the selected case studies.

In [99] the authors built a complex ML structure based on assembling neural networks and fuzzy inference combined with different techniques for selecting the input variables (genetic algorithms, simulated annealing, imperialist competitive algorithm, and differential evolution). They found that the meta-heuristic approach using differential evolution technique provided the best Root Mean Squared Error (RMSE) when feeding the model with variables related to land use, vegetation, land morphology, and lithology, among others.

Similarly, [100] investigated the capacity of a Random Forest algorithm to predict the daily discharge using the meteorological and hydrology features. They compared the output from the data-driven algorithm with a partial, physical approach. . The two methods provided results with comparable accuracy, which led the authors to conclude that the use of the data-driven model is advisable because the random forest was fed only with precipitation records, while the physical model required a range of inputs and parameters.

Selecting the correct set of physically related inputs and outputs is the key point for properly defining data-driven algorithms with physical criteria. Some works point to a strategy for the optimal selection of inputs. The authors in [101] presented a complex method that combined physical criteria with a genetic algorithm for selecting the input variables and decomposition technique for preprocessing the data of the selected variables. The output of the previous process was used to feed both neural networks and extreme ML. The method suggested precipitation, rainfall, and lagged streamflow should provide the best performance in predicting future monthly streamflow. Extreme ML was revealed as the most accurate. The work [102] proved that using a simple long short-term memory (LSTM) network for the ensemble modeling of 531 basins with a mass conservation law approach achieved better performance than hydrological models that were calibrated individually. They used a simple physical relationship feeding the model with meteorological and land-related variables (aridity, elevation, forests) to predict streamflow, even presenting the features' importance in different catchments.

c: DATA-DRIVEN MODELS BASED ON CONCEPTUAL APPROACHES

In this case, researchers use data-driven models based not on pure physical relationships, but rather on conceptual theories and models as presented in (1), similar to what is done in hydrological distributed models. Most authors define conceptual approaches on a mass conservation law relating the recorded flow discharge in several tributaries to predict the observed flow in the main river. Following this line [103] achieved reasonable performance for predicting daily streamflow in a river basin in Illinois using artificial neural networks aided by a wavelet transformation. Similarly, [104] used monotone composite quantile regression neural network coupled with a (local scope) conceptual model called Xinanjiang hydrological model for short-term (three hours) flood forecasting. They feed the model with both rainfall information and runoff from tributary catchments, which suggests an intermediate approach between conceptual and physically based. The developed model performed well when predicting extreme events which, given the time scale of three hours, is a result of interest for hydrologists.

IV. DISCUSSION AND OPEN CHALLENGES

Based on the dissertation of the previous sections, we can conclude that hydrology is a holistic discipline, where the combination of individual processes, analyzed in isolation, does not offer a reliable picture of the whole system, as interactions cannot be neglected [105]. Hydrologists recognize that to advance in the hydrological sciences, rather than new modeling approaches there is a need for new measurement techniques and equipment, as well as more field measurements [11]. In the following of this section, we discuss the many open challenges that call for novel modeling approaches built upon high-quality data.

A. MONITORING OF EVENT-BASED HYDROLOGICAL PROCESSES

Though highly accurate experimental techniques are nowadays available, hydrologists still claim high-quality data as the best alternative for improving the understanding of hydrological processes [12]. However, high-quality experimental data do not necessarily yield highly detailed process measurement or characterization, since complex and interlinked processes as those involved in hydrological systems require not only accurate measurements, but also appropriate methodologies to place the sensors, tune their parameters, and analyze their readings. In certain cases, a detailed measurement of a given process does not provide better results than other inaccurate but correctly distributed measurements. Examples of this issue are given in [106] and [107] for rainfall estimation.

In this respect, hydrologists can benefit from the broad literature on monitoring strategies, including theoretical studies [108], [109], [110], [111], [112] as well solutions developed in various disciplines [113], [114], [115], [116], [117], [118], [119]. With regard to hydrology specifically, the problem of defining optimal monitoring networks has been studied in relation to certain hydrological processes in isolation [120], [121], while the whole hydrological system has not yet been considered.

It is worth noting that the monitoring problem does not end with the design of the monitoring network, but also involves the sampling strategy, which must adapt to the dynamics of the hydrological processes. The system is indeed complex, with extreme spatial and temporal variability, so the swift detection of trends can help accurately track the system's evolution.

To achieve this goal, strategies for hydrology's processes monitoring must shift into novel approaches based on the optimal combination of different measurement techniques (remote, on-site, unmanned) and capable of dynamically (spatially and temporally) adapting measurement strategies in view of the evolution of the processes involved. This approach must be based on a comprehensive theoretical model with a clear target (i.e., forecasting water depth in a course during and after a storm event, forecasting aquifers recharge as a result of increasing water reservoirs capacity, etc.) to decide what has to be measured, where and with what latency the data from what sensor must be collected and processed to obtain the best information and, thus, the optimal input for modeling and forecasting hydrology.

The problem hence consists of selecting the combination of (mobile and/or fixed) sensors providing the most accurate information on the process evolution at each time. This problem has been previously addressed in the literature from different perspectives, not directly applied to hydrology as, for example, in [122], [123], [124], [125], and [126]. Although the specificity of the hydrological context makes standard monitoring solutions ineffective or underperforming, yet the state-of-the-art methodologies developed for designing and managing mobile sensor networks can inspire solutions for operating hydrological monitoring networks. Some works (e.g., [127]) focus on the adaptive sensor placement in view of the system evolution, which could be of interest for the hydrological problem since they define a recursive algorithm for adaptive sensor location based on real-time observation of environmental variables.

Mobile sensors (mainly unmanned aerial vehicles) have recently gained traction as a valuable complement for on-ground sensors for hydrological applications [41]. However, they are not sufficient to fill the gap towards practical real-time remote monitoring of hydrological processes because of their high investment and operation costs, and logistics issues, such as the practicality (or even feasibility) of rapidly deploying large fleets of

UAVs to remote locations to monitor short-term storm events.

Therefore, the problem of *which, where, and when* sensing measurements shall be taken is still largely unsolved in the hydrological sector, and methodologies are needed to the dynamically adjust the sampling frequency and even the location of detection nodes to better predict the occurrence of extreme events.

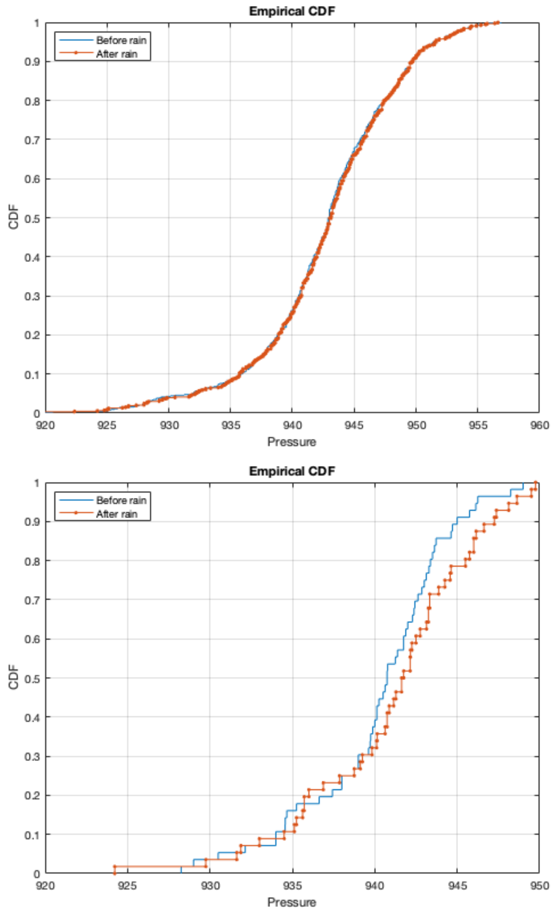


FIGURE 5. CDF of the atmospheric pressure before (blue line) and after (red dotted line) rainfall. Upper graph: all rainfalls. Lower graph: only rainfalls with overall precipitation volume larger than 5 mm.

B. MODELING EVENT-BASED HYDROLOGICAL PROCESSES

In spite of the amount and variety of algorithms used to model and predict hydrology events, there are still challenges that have not been successfully addressed by previous studies. In particular, adequate monitoring of event-based hydrological processes requires accurate prediction of the time evolution of both rainfall and runoff and flood transients, which are still open problems as better explained in the following.

1) MODELING STORMS

For what concerns the modeling of storm events, the focus should be on hyetographs (i.e., the intensity of rainfall over time) rather than the cumulative precipitation in long-time intervals, as is mostly done today. In fact, monthly or daily precipitation forecasts are important but inadequate for forecasting the evolution of hydrological systems under storms. Event-based hydrology needs finer temporal detail.

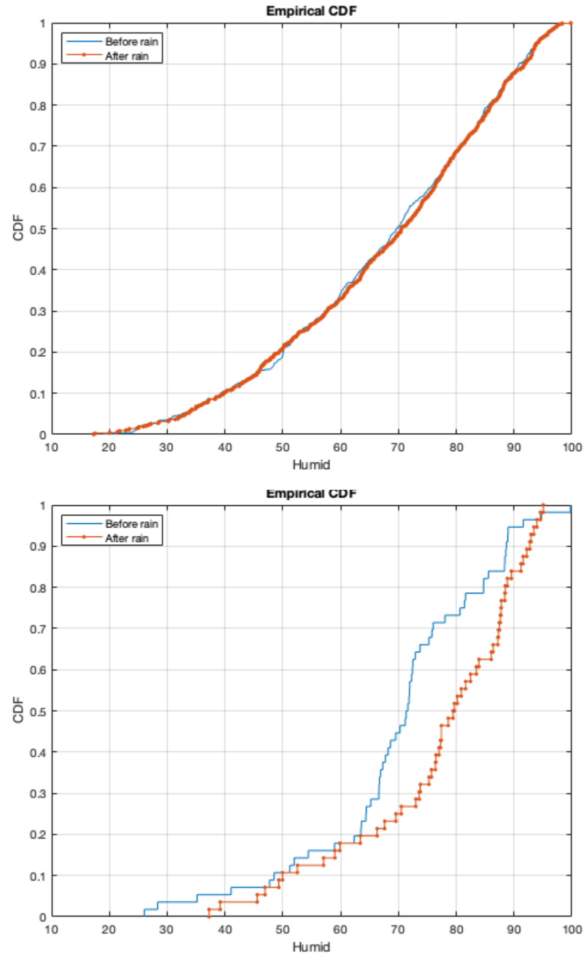


FIGURE 6. CDF of the humidity before and after rainfall. Upper graph: all rainfalls. Lower graph: only rainfalls with overall precipitation volume larger than 5 mm.

Despite its apparent obviousness, this requirement is often disregarded in practice. Up to hourly precipitation time steps might be enough for large catchments, though they are not suitable to capture heavy shower events that might be critical in certain cases. Some works have previously attacked this question using neural networks [128], [129], [130], logistic regression [131], long-short term memory or support vector regression [132]. Yet, large databases with sub-hourly rainfall measurements are not common, which hinders the development of algorithms that can exploit such finer-grained time series.

However, short-term rainfall prediction is of paramount importance, and efficient algorithms with this aim must be prioritized. Physical relationships for precipitation suggest that autoregressive models could work properly for short-term prediction of storms' evolution since the amount of water available to rain at any time depends on the

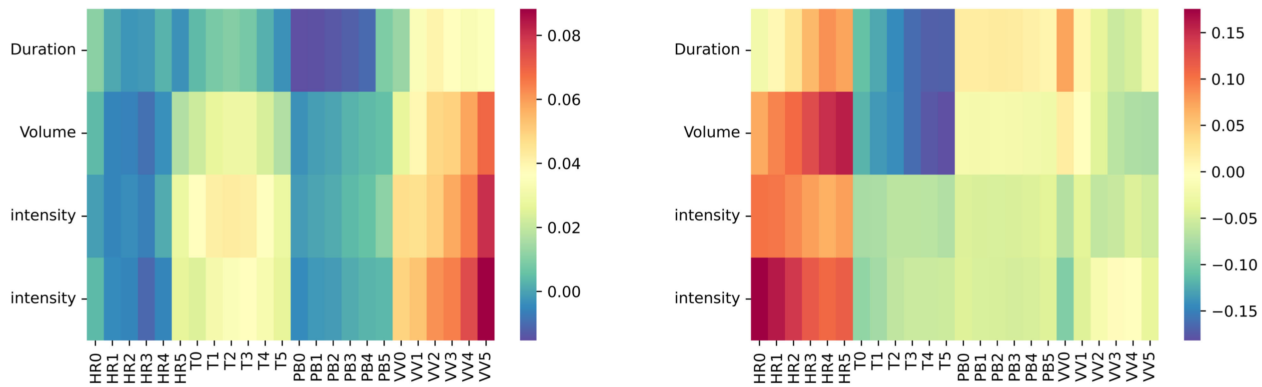


FIGURE 7. Linear correlation coefficients between storm-related variables (y-axis) and relative humidity -HR-, temperature -T-, atmospheric pressure -PB- and wind velocity -VV-, one, two, three, four and five hours before the storm starts and during the first hour (ticks of the x-axis refer to the five hourly measurements of each parameter).

previously rained water. Therefore, shaping data-driven algorithms with physical criteria, and feeding the models with variables triggering precipitation (atmospheric pressure, relative humidity for example), will likely help improve forecasting accuracy. While autoregressive approaches often use ARIMA [133], the methods presented in works exploiting physical causality use a wider range of techniques, as highlighted by the references included in the previous paragraph.

Although the previous discussion argues for more in-depth physical relationships, the correlation of different environmental and atmospheric variables with the intensity and duration of precipitation events is also sometimes elusive. For example, Fig. 5 shows the cumulative distribution function (CDF) of the atmospheric pressure immediately before (blue line) and after (red line with dot markers) a rainfall. The upper graph is obtained by considering all rainfall events from January 1^{st}, 2019, to September 30th, 2022, collected in Madrid's catchment. The lower graph, instead, was acquired by considering only storm events with an aggregate volume of precipitation during the storm larger than 5 mm, which is equal to the mean volume of water dropped during the storms in the considered period, augmented by one standard deviation of the precipitation volume per storm (in other words, we focused on storms that delivered more water than the average). Fig. 6 reports similar results for the humidity registered immediately before and after rainfall. In both cases, we can notice that the CDF curves are almost perfectly overlapping when we consider all the rainfall events, including those with very light rain, which are the majority. In this case, atmospheric pressure and humidity are basically unaffected by the rainfall event. However, if we limit our attention to events with relatively large precipitation volumes we can clearly see that the curves separate, indicating that pressure and humidity typically increase as a consequence of an intense rainfall event. This means that there is no clear pattern to guide the modeling strategy and the relationship between storm-related metrics of interest (rained volume, storm duration, and either maximum or average rainfall intensity) and the set of variables apparently triggering storms (relative humidity, temperature, atmospheric pressure, or wind velocity) remains rather elusive.

Evidence of this problem is provided by Fig. 7, which displays the linear correlation coefficients between the aforementioned storm-related metrics and the relative humidity, temperature, atmospheric pressure, and wind velocity recorded one, two, three, four, and five hours before the storm starts (and during the first hour of the storm, as well) with data from six different weather stations located in Madrid (hourly data between January 1^{st}, 2019, to September 30th, 2022). For the heat map, we can see that no strong linear relationship is observed in any of the cases. In addition, from the scatterplots shown in Fig. 8 there is no clear trace of even non-linear relationships between the processes. therefore, the physical causality between the previously mentioned variables remains hidden behind the records. It implies that data-driven approaches must be shaped to capture the hidden patterns so hydrology can leverage the prediction of the potential evolution of storm events from the previous measurements of atmospheric variables.

2) MODELING RAINFALL-RUNOFF PROCESSES

The particular nature of the rainfall-runoff process makes any data-driven algorithm (particularly black-box approaches) apparently well suited for modeling them. As a matter of fact, the problem has been addressed by researchers using neural networks with different architectures [104], [134], [135], [136], [137], random forest [100], [138], SVMs [95], [96] among many others. However, previous attempts at addressing hydrological processes with data-driven models either completely neglect or consider rough and simplified versions of the underlying hydrological models. They mostly exploit existing datasets to build predictive models upon scarcely connected variables without structured physically-based causal relationships. Similarly, input data

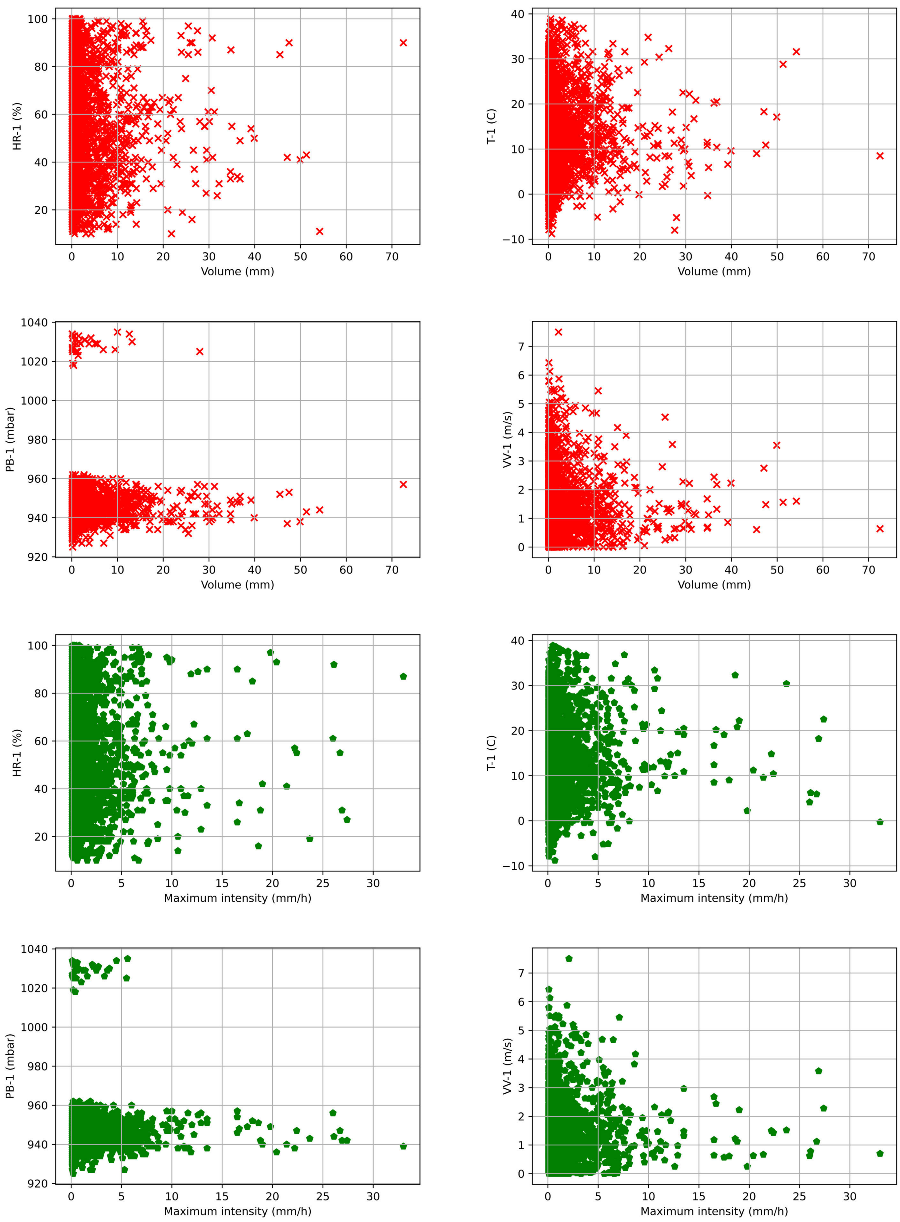


FIGURE 8. Scatterplots of both rained volume and relative humidity -HR-, temperature -T-, atmospheric pressure -PB- and wind velocity -VV- one hour before the storm starts.

are frequently pre-processed following data-driven requirements but ignoring potential physical relationships among variables recommending specific input data structures. In this frame, the statistical significance becomes the main target, leading to the selection of the model with the best significance metrics that, however, does not necessarily provide the most useful output. For example, a model systematically predicting the absence of runoff will surely provide high-performance metrics, considering that the vast majority of storm events do not generate runoff. However, if the model fails when predicting (very rare) extreme events, it cannot be a useful tool. Similarly, to improve the understanding of hydrological processes, we should deviate from the classical approach based on picking physically-based models for the selected hydrological processes, setting the optimal data-driven algorithm (type and architecture) accordingly, and gathering the data to feed that specific algorithm. Instead, we should merge physically-based and data-driven approaches, so that the second can learn from the outputs of the first one, which in turn can be improved based on the insights provided by the data-driven model. For example, using Philip's model we can estimate the occurrence of runoff for a certain catchment, given a set of input (measurable) variables (e.g., pressure, relative humidity, temperature). From this (synthetic, but theoretically grounded) output data, we can extract the events generating runoff, thus retrospectively filtering the input data to focus only on those that are actually related to the occurrence of the event of interest (i.e., runoff). Then, we can feed these selected data to ML algorithms that are trained to predict the runoff occurrence from the set of variables triggering the storms.

This approach can be adopted not only to predict the evolution of a given target variable but also to shed light on many complex parameters involved in physical models. For example, the theoretical model can be re-applied with the same input data, but different choices of other system parameters (e.g., soil hydraulic properties), and the ML approach can be used to infer the importance and impact of the single parameters. Models' parametrization is indeed an open problem [139] addressed by many authors with different approaches [140], [141]. Physical models are often overparametrized to better fit observations (see for example soil conductivity or water retention curves). On many occasions, the accurate determination of all the parameters involved represents a very challenging task, and reducing the number of parameters would be extremely advantageous. While physical approaches for models' parametrization seek to advance by improving the comprehension of the underlying physical principles, data-driven approaches have also been used to determine the parameters, mostly based on optimization models [142]. Addressing the parametrization through data-driven models can also shed light on several blurred hydrological concepts, for example, the time of concentration (discussed in [143]), the transit time (also analyzed in [144]) or even the flood routing.

Previous attempts in using ML algorithm for parameters' calibration simply seek to provide efficient tools for finding the values of the model's parameters in their actual configuration. With this aim, researchers have leveraged the ML algorithms as optimization tools using mainly deep learning [145], [146], [147] and other algorithms as regression trees [148]. However, ML algorithms should pave the way for streamlining models, and shaping supervised data-driven models in view of physical ones can help deduce a better parameters' structure identifying self-correlations, redundant information, or patterns between dependent and independent variables that were previously unknown. Reference [149] tangentially discussed the ability of deep learning for addressing such tasks but it will undoubtedly be worth delving into this research line.

3) DEALING WITH SPATIAL VARIABILITY

The literature has widely recognized the limitations that extreme spatial and temporal variability of precipitation poses to gathering high-quality data. For dealing with spatial variability, hydrology has traditionally used deterministic approaches based on, for example, Kriging techniques or Thiessen polygons (as, for example, that presented by [150]). However, those deterministic techniques are far from being accurate, since the results depend on the interpolation method that is used, the density and location of the measurement stations, and the sampling frequency of the rainfall. To exemplify the concept, we report in Fig. 9 the results given by three different interpolation techniques used to obtain a continuous map of precipitation from the same set of records, collected by six weather stations located in Madrid. As can be noticed, the estimate of precipitation intensity over the catchment is very different for the three cases, as is the estimate of the total amount of rainfall over the drainage area.

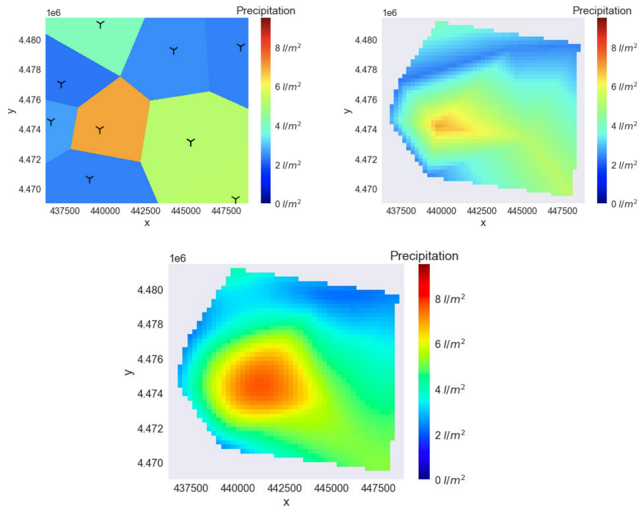


FIGURE 9. Example of the spatial distribution of rainfall intensity over Madrid catchment obtained through Thiessen polygons, barycentric, and Clough-Tocher interpolation of a discrete number of measurement stations.

Moreover, deterministic approaches are clearly unsuitable for dealing with the stochastic evolution of storms across a complex catchment. Catchments are often larger than that presented in Fig. 4 (and its spatial variability drafted in Tab. 2) with a greater variety of land uses and morphology. As an example, Fig. 10 shows some images of the morphology and land uses of a medium-sized river basin located at the South-est of Madrid city. Spatial variability must therefore be considered together with temporal variability, as advocated when discussing the need for adaptive monitoring strategies. Some attempts in this direction have been made for certain hydrological processes, such as precipitation and others, in [120] and [151]. As usual, authors spread over a wide range of ML algorithms from SVMs to convolutional neural networks through long-short memory networks, multi-layer perceptrons, extreme ML, or combinations of those and other architectures. For example, [152] defined an integrated extreme learning method predicting the observed streamflow from observations from 15 tributary measurement stations and compared the results with those obtained with a geomorphology-based extreme learning method. They observed that the integrated learning method is capable of exploring the spatial variation of the rainfall–runoff process without requiring the physical characteristics of sub-basins. The paper [151] compared the ability of different ML algorithms for efficient merging of satellite and on-site

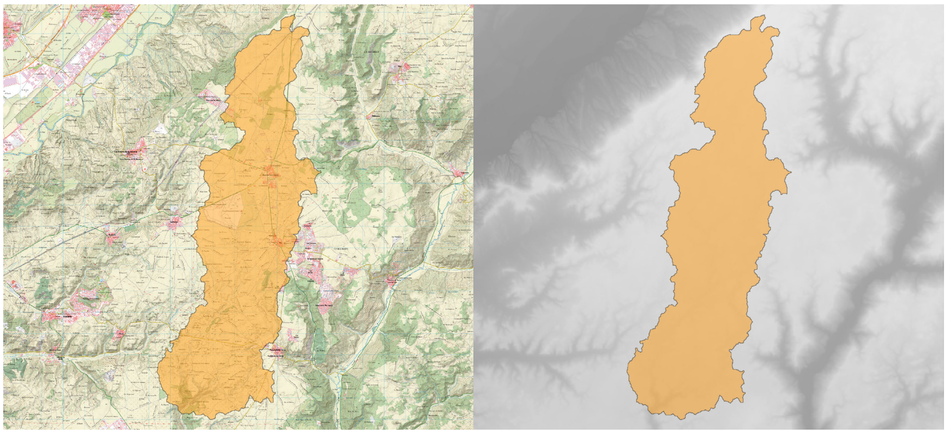
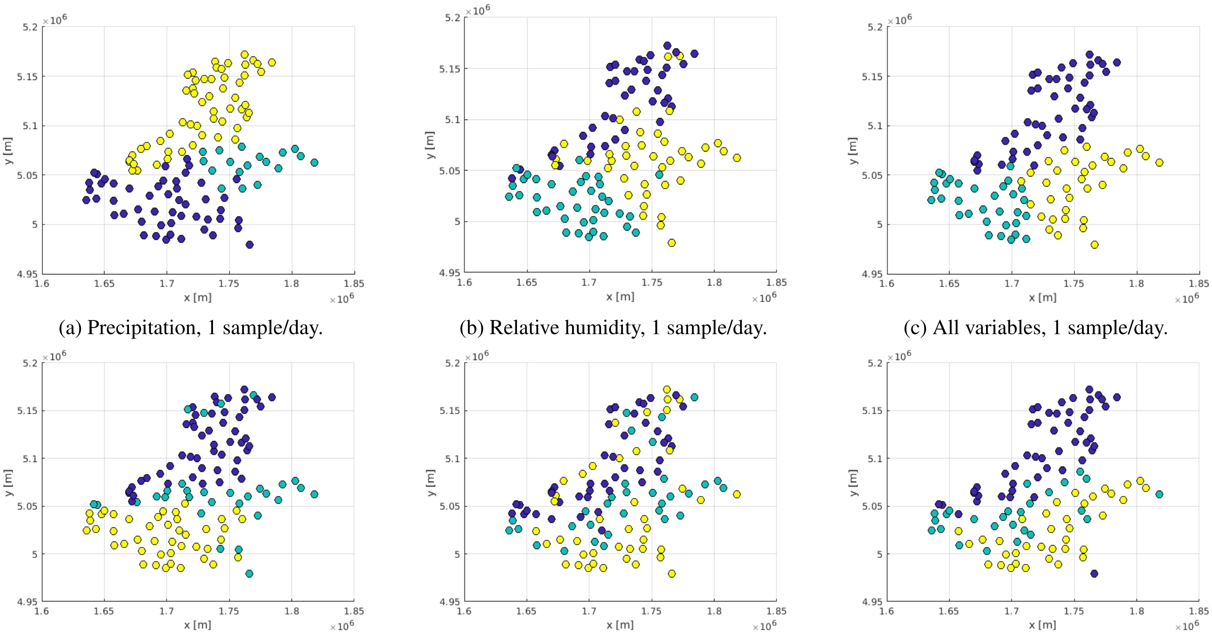


FIGURE 10. Example of a medium size (5000000 m^2) catchment in Madrid.



(a) Precipitation, 1 sample/day.
(b) Relative humidity, 1 sample/day.
(c) All variables, 1 sample/day.
(d) Precipitation, 4 samples/hour.
(e) Relative humidity, 4 samples/hour.
(f) All variables, 4 samples/hour.

FIGURE 11. Clustering of meteorological stations in the Veneto region (IT) using as features different combinations of environmental variables (precipitation, relative humidity, air temperature, solar radiation) and sampling periods (24 h and 15'). Clusters are identified by different colors.

precipitation records considering spatiotemporal variability. They found an ensemble approach using convolutional neural networks with long-short memory networks provided better performance.

Scientists are deeply concerned about the issues the spatiotemporal variability poses to hydrological modeling. From rainfall spatial variability and its effect on runoff generation [152] to soil water content estimation [153], the literature has mainly addressed spatial variability focusing on *individual* processes instead of the whole hydrological system. Hence, there is a clear gap in addressing such variability for the hydrological system as a whole and merging data-driven approaches and conceptual models. Defining homogeneous areas in terms of hydrological behavior could also help address the spatial variability and define optimal monitoring strategies. This line has been explored, using criteria based on either physical models (see for example [154] that used the coefficient of variation of a sort of normalized discharged flow for different return periods) or ML ones (see [155] using hierarchical clustering of principal components for grouping watershed using climate, geological, topographical, and land-cover data). However, as discussed when reporting about the relationship between the storm-related metrics and the variables apparently triggering the storms, the clustering strategy may suffer from the *a priori* choice of the measured variables and sampling frequency. As an example, we clusterized meteorological stations in the Veneto region, in the north of Italy, based on the similarities between time series of certain variables collected at a given sampling rate. We used data gathered by the weather stations in year 2017, with the purpose of identifying stations with similar measurements and, in turn, areas having homogeneous weather characteristics. Fig. 11 shows the results for different choices of the input variables, fixing the number of clusters to 3 and using the K-medoids clustering algorithm. The inputs to the clustering algorithm were vectors containing the timeseries of precipitation (figures (a)), relative humidity (figures (b)), and all variables (figures (c)), aggregated over a time period of 1 day (top row) or only 15 minutes (lower row), as indicated in the corresponding captions. The dots in the figures represent the geographical location of the weather stations, and the color indicates the cluster to which they belong. Comparing the figures it is apparent that, although the clusters roughly correspond to the three geographical regions (mountain, plain, Venetian lagoon/coast), the actual cluster members differ depending on the choice of the input variables. For example, while the clusters obtained considering only the precipitation (Figs. 11a, 11d) are geographically compact, those obtained with the relative humidity (Figs. 11b, 11e) are more spread. However, certain stations are always clustered together, indicating a stronger correlation between their measurements and, consequently, possibly lower mutual information. In contrast, the stations that are associated with different clusters depending on the input variables are potentially more informative.

Comparing the upper and lower rows of graphs, we can notice that also the sampling frequency plays a role in the cluster determination. Interestingly, finer temporal sampling produces clusters that tend to be more horizontally distributed (and more spread) than those obtained from daily aggregate measurements. It is unclear whether this result has physical significance or derives from numerical artifacts (e.g., the temporal shifting of the measurements due to the time variability of the hydrological processes), but it clearly exemplifies the importance of proper data selection when applying data-drive inspection methods.

V. CONCLUSION AND OPEN QUESTIONS

Current ICT solutions, such as sensor networks and artificial intelligence, can be useful in modeling and adjusting parameters based on measured data. However, as explained in this article, modeling and predicting water flows and distribution means considering a complex system that is influenced by many interrelated variables and elements (e.g., land characteristics and use, and weather conditions). Therefore, to address these problems, it is necessary to better understand several aspects related to the observation, processing, and interpretation of hydrological processes. Scientists from the ICT and computing disciplines are then required to propose novel approaches, addressing the following open research questions.

A. MONITORING AND SENSING

efficient and adaptive monitoring strategies for event-based hydrological processes have to be provided. The optimal strategy will have to efficiently exploit on-site and remote technologies. Information collected by *in situ* sensors can be complemented with satellite maps and aerial images taken by drones to obtain more precise geographical and topographical information, which can be used to fine-tune the parameters of hydrologic models. The extreme spatial and temporal variability will have to be addressed in the search for the optimal monitoring strategy. A number of different challenges should be addressed by the scientist: How many gauging stations are needed to correctly estimate the amount of rained water in catchments? Which type of sensors should be provided to such stations? What should be the sampling period for the different environmental variables? How the monitoring strategy can be timely adapted in view of the system's evolution? How can we monitor large and remote geographical areas with energy-harvesting technologies, combined with energy-efficient data collection and transmission protocols?

B. MODELING AND PREDICTING

Data-driven models shaped by physical and conceptual criteria are expected to provide important insights for modeling and predicting event-based hydrological processes. Suitable data-driven algorithms can be used to extract information from the sensing data and improve the accuracy of process-based hydrology models, thus offering more reliable predictions of the water fluxes. Methodologies coming from the complex networks domain can be exploited to model the interactions of the different water subsystems and determine the spatiotemporal dynamics of water flows across large areas. Similarly, models for refining and streamlining physical theories reducing the current over-parametrization are needed to help hydrologists gain a better understanding of complex physical theories. In conclusion, a novel paradigm emerging from a radical shift in event-based hydrology modeling is required to take a step forward in hydrological systems, and sensing, data processing and inference techniques are instrumental in fulfilling this ambitious objective.

ACKNOWLEDGMENT

The authors would like to thank Dr. Martina Capuzzo for her contribution to the production of part of the results presented in this article. In Spain, this article belongs to PCI2020-120694-2 Project funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR.

REFERENCES

[1] C. W. Thornthwaite and B. Holzman, "The determination of evaporation from land and water surfaces," *Monthly Weather Rev.*, vol. 67, no. 1, pp. 4–11, Jan. 1939.
[2] H. Darcy, *Les Fontaines Publiques De La Ville De Dijon: Exposition Et Application*. Paris, France: Victor Dalmont, Libraire des Corps imperiaux des ponts et chaussées et des mines, 1856.

[3] J. R. Philip, "The theory of infiltration: 4. Sorptivity and algebraic infiltration equations," *Soil Sci. J.*, vol. 84, pp. 257–264, Sep. 1957.
[4] L. A. Richards, "Capillary conduction of liquids through porous mediums," *Physics*, vol. 1, no. 5, pp. 318–333, Nov. 1931.
[5] W. H. Green and G. A. Ampt, "Studies in soil physic, I: The flow of air and water through soils," *J. Agric. Sci.*, vol. 4, pp. 1–24, 1911, doi: 10.1017/S0021859600001441.
[6] B. Saint-Venant, "Theory of unsteady water flow, with application to river floods and to propagation of tides in river channels," *Fr. Acad. Sci.*, vol. 73, pp. 148–154, 1871.
[7] A. N. Kostiakov, "On the dynamics of the co-efficient of water percolation in soils," in *Proc. 6th Cong. Int. Soil Sci.*, 1932, pp. 15–21.
[8] R. E. Horton, "The role of infiltration in the hydrologic cycle," *Trans. Amer. Geophys. Union*, vol. 14, no. 1, pp. 446–466, 1933.
[9] R. E. Horton, "Analysis of runoff-plat experiments with varying infiltration-capacity," *Trans. Amer. Geophys. Union*, vol. 2, no. 44, pp. 693–711, 1933.
[10] F. Tauro, J. Selker, N. Van De Giesen, T. Abrate, R. Uijlenhoet, M. Porfiri, S. Manfreda, K. Caylor, T. Moramarco, J. Benveniste, and G. Ciraolo, "Measurements and observations in the XXI century (MOXXI): Innovation and multi-disciplinarity to sense the hydrological cycle," *Hydrological Sci. J.*, vol. 63, no. 2, pp. 169–196, Jan. 2018.
[11] G. Blöschl, M. F. Bierkens, A. Chambel, C. Cudennec, G. Destouni, A. Fiori, J. W. Kirchner, J. J. McDonnell, H. H. Savenije, M. Sivapalan, and C. Stumpp, "Twenty-three unsolved problems in hydrology (UPH)—A community perspective," *Hydrolog. Sci. J.*, vol. 64, no. 1, pp. 1141–1158, 2017.
[12] T. Blume, I. van Meerveld, and M. Weiler, "The role of experimental work in hydrological sciences-insights from a community survey," *Hydrolog. Sci. J.*, vol. 62, no. 3, pp. 334–337, 2017.
[13] M. T. van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," *Soil Sci. Soc. Amer. J.*, vol. 44, no. 5, pp. 892–898, Sep. 1980.
[14] Y. Mualem, "A new model for predicting the hydraulic conductivity of unsaturated porous media," *Water Resour. Res.*, vol. 12, no. 3, pp. 513–522, Jun. 1976.
[15] R. F. Carsel and R. S. Parrish, "Developing joint probability distributions of soil water retention characteristics," *Water Resour. Res.*, vol. 24, pp. 755–769, May 1988.
[16] L. Brocca, L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani, "Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data," *J. Geophys. Res., Atmos.*, vol. 119, no. 9, pp. 5128–5141, May 2014.
[17] C. Massari, W. Crow, and L. Brocca, "An assessment of the performance of global rainfall estimates without ground-based observations," *Hydrol. Earth Syst. Sci.*, vol. 21, no. 9, pp. 4347–4361, Sep. 2017.
[18] S. Manfreda, K. K. Caylor, and S. P. Good, "An ecohydrological framework to explain shifts in vegetation organization across climatological gradients," *Ecohydrology*, vol. 10, no. 3, p. e1809, Apr. 2017.
[19] L. Zhu, H. Wang, C. Tong, W. Liu, and B. Du, "Evaluation of ESA active, passive and combined soil moisture products using upscaled ground measurements," *Sensors*, vol. 19, no. 12, p. 2718, Jun. 2019.
[20] X. Meng, K. Mao, F. Meng, X. Shen, T. Xu, and M. Cao, "Long-term spatiotemporal variations in soil moisture in North East China based on 1-km resolution downscaled passive microwave soil moisture products," *Sensors*, vol. 19, no. 16, p. 3527, Aug. 2019.
[21] M. Farokhi, F. Faridani, R. Lasaponara, H. Ansari, and A. Faridhosseini, "Enhanced estimation of root zone soil moisture at 1 km resolution using SMAR model and MODIS-based downscaled AMSR2 soil moisture data," *Sensors*, vol. 21, no. 15, p. 5211, Jul. 2021.
[22] K. Chen, X. Cao, F. Shen, and Y. Ge, "An improved method of soil moisture retrieval using multi-frequency SNR data," *Remote Sens.*, vol. 13, no. 18, p. 3725, Sep. 2021.
[23] E. P. Glenn, C. M. U. Neale, D. J. Hunsaker, and P. L. Nagler, "Vegetation index-based crop coefficients to estimate evapotranspiration by remote sensing in agricultural and natural ecosystems," *Hydrol. Process.*, vol. 25, pp. 450–462, Dec. 2011.
[24] W. G. M. Bastiaanssen, M. Menenti, R. A. Feddes, and A. A. M. Holtslag, "A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation," *J. Hydrol.*, vols. 212–213, pp. 198–212, Dec. 1998.
[25] F. Castellví, C. Cammalleri, G. Ciraolo, A. Maltese, and F. Rossi, "Daytime sensible heat flux estimation over heterogeneous surfaces using multitemporal land-surface temperature observations," *Water Resour. Res.*, vol. 52, no. 5, pp. 3457–3476, May 2016.
[26] G. J.-P. Schumann and A. Domeneghetti, "Exploiting the proliferation of current and future satellite observations of rivers," *Hydrological Processes*, vol. 30, no. 16, pp. 2891–2896, Jul. 2016.
[27] A. Domeneghetti, A. Castellarin, A. Tarpanelli, and T. Moramarco, "Investigating the uncertainty of satellite altimetry products for hydrodynamic modelling," *Hydrol. Process.*, vol. 29, pp. 4908–4918, Nov. 2015.
[28] R. Schneider, P. N. Godiksen, H. Villadsen, H. Madsen, and P. Bauer-Gottwein, "Application of CryoSat-2 altimetry data for river analysis and modelling," *Hydrol. Earth Syst. Sci.*, vol. 21, pp. 751–764, Feb. 2019.
[29] M. J. Tourian, A. Tarpanelli, O. Elmi, T. Qin, L. Brocca, T. Moramarco, and N. Sneeuw, "Spatiotemporal densification of river water level time series by multimission satellite altimetry," *Water Resour. Res.*, vol. 52, no. 2, pp. 1140–1159, Feb. 2016.
[30] M. Durand, K. M. Andreadis, D. E. Alsdorf, D. P. Lettenmaier, D. Moller, and M. Wilson, "Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model," *Geophys. Res. Lett.*, vol. 35, no. 20, p. 241, 2008.
[31] C. A. Schlosser and P. R. Houser, "Assessing a satellite-era perspective of the global water cycle," *J. Climate*, vol. 20, no. 7, pp. 1316–1338, Apr. 2007.
[32] V. Levizzani and E. Cattani, "Satellite remote sensing of precipitation and the terrestrial water cycle in a changing climate," *Remote Sens.*, vol. 11, no. 19, p. 2301, Oct. 2019.
[33] C. Kidd, G. Huffman, V. Maggioni, P. Chambon, and R. Oki, "The global satellite precipitation constellation: Current status and future requirements," *Bull. Amer. Meteorol. Soc.*, vol. 12, no. 1, pp. 1844–1861, 2021.
[34] J. Fang, W. Yang, Y. Luan, J. Du, A. Lin, and L. Zhao, "Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China," *Atmos. Res.*, vol. 223, pp. 24–38, Jul. 2019.
[35] S. Prakash, A. K. Mitra, D. S. Pai, and A. AghaKouchak, "From TRMM to GPM: How well can heavy rainfall be detected from space?" *Adv. Water Resour.*, vol. 88, pp. 1–7, Feb. 2016.
[36] H. Chen, B. Yong, Y. Shen, J. Liu, Y. Hong, and J. Zhang, "Comparison analysis of six purely satellite-derived global precipitation estimates," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124376.
[37] H. R. Bogena, J. A. Huisman, A. Güntner, C. Hubner, J. Kusche, F. Jonard, S. Vey, and H. Vereecken, "Emerging methods for noninvasive sensing of soil moisture dynamics from field to catchment scale: A review," *WIREs Water*, vol. 2, no. 6, pp. 635–647, Nov. 2015.
[38] J. A. Richards, *Remote Sensing Digital Image Analysis*, vol. 5, New York, NY, USA: Springer, 2022.
[39] C. Xu, Y. Wang, H. Fu, and J. Yang, "Comprehensive analysis for long-term hydrological simulation by deep learning techniques and remote sensing," *Frontiers Earth Sci.*, vol. 10, p. 546, Apr. 2022.
[40] M. Avand, H. Moradi, and M. R. Lasboyee, "Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability," *J. Hydrol.*, vol. 595, Apr. 2021, Art. no. 125663.
[41] M. Vélez-Nicolás, S. García-López, L. Barbero, V. Ruiz-Ortiz, and Á. Sánchez-Bellón, "Applications of unmanned aerial systems (UASs) in hydrology: A review," *Remote Sens.*, vol. 13, no. 7, p. 1359, Apr. 2021.
[42] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.
[43] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 1, pp. 19–38, Mar. 2022.
[44] Z. Salahuddin, H. C. Woodruff, A. Chatterjee, and P. Lambin, "Transparency of deep neural networks for medical image analysis: A review of interpretability methods," *Comput. Biol. Med.*, vol. 140, Jan. 2022, Art. no. 105111.
[45] C. H. Chen, *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 2015.
[46] D. Crevier and R. Lepage, "Knowledge-based image understanding systems: A survey," *Comput. Vis. Image Understand.*, vol. 67, no. 2, pp. 161–185, Aug. 1997.
[47] U. Iqbal, P. Perez, W. Li, and J. Barthelemy, "How computer vision can facilitate flood management: A systematic review," *Int. J. Disaster Risk Reduction*, vol. 53, Feb. 2021, Art. no. 102030.

[48] M. Zreda, W. J. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz, and R. Rosolem, ''COSMOS: The cosmic-ray soil moisture observing system,'' *Hydrol. Earth Syst. Sci.*, vol. 16, pp. 479–499, Nov. 2012.

[49] K. M. Larson, E. E. Small, E. D. Gutmann, A. L. Bilich, J. J. Braun, and V. U. Zavorotny, ''Use of GPS receivers as a soil moisture network for water cycle studies,'' *Geophys. Res. Lett.*, vol. 35, no. 24, pp. 142–149, 2008.

[50] G. Calamita, A. Perrone, L. Brocca, B. Onorati, and S. Manfreda, ''Field test of a multi-frequency electromagnetic induction sensor for soil moisture monitoring in southern Italy test sites,'' *J. Hydrol.*, vol. 529, pp. 316–329, Oct. 2015.

[51] N. Romano, ''Soil moisture at local scale: Measurements and simulations,'' *J. Hydrol.*, vol. 516, pp. 2–6, Aug. 2014.

[52] Y. Rothfuss, H. Vereecken, and N. Brüggemann, ''Monitoring water stable isotopic composition in soils using gas-permeable tubing and infrared laser absorption spectroscopy,'' *Water Resour. Res.*, vol. 49, no. 6, pp. 3747–3755, 2013.

[53] T. H. M. Volkmann and M. Weiler, ''Continual in situ monitoring of pore water stable isotopes in the subsurface,'' *Hydrol. Earth Syst. Sci.*, vol. 18, no. 5, pp. 1819–1833, May 2014.

[54] C. Sayde, C. Gregory, M. Gil-Rodriguez, N. Tufillaro, S. Tyler, N. van de Giesen, M. English, R. Cuenca, and J. S. Selker, ''Feasibility of soil moisture monitoring with heated fiber optics,'' *Water Resour. Res.*, vol. 46, no. 6, Jun. 2010.

[55] Y. Nihei and A. Kimizu, ''A new monitoring system for river discharge with horizontal acoustic Doppler current profiler measurements and river flow simulation,'' *Water Resour. Res.*, vol. 44, no. 4, Apr. 2008.

[56] E. Garel and D. D'Alimonte, ''Continuous river discharge monitoring with bottom-mounted current profilers at narrow tidal estuaries,'' *Continental Shelf Res.*, vol. 133, pp. 1–12, Feb. 2017.

[57] A. J. F. Hoitink, F. A. Buschman, and B. Vermeulen, ''Continuous measurements of discharge from a horizontal acoustic Doppler current profiler in a tidal river,'' *Water Resour. Res.*, vol. 45, no. 11, p. 1146, Nov. 2009.

[58] M. Perumal, T. Moramarco, B. Sahoo, and S. Barbetta, ''A methodology for discharge estimation and rating curve development at ungauged river sites,'' *Water Resour. Res.*, vol. 43, no. 2, p. 2412, Feb. 2007.

[59] A. Crabit, F. Colin, J. S. Bailly, H. Ayroles, and F. Garnier, ''Soft water level sensors for characterizing the hydrological behaviour of agricultural catchments,'' *Sensors*, vol. 11, no. 5, pp. 4656–4673, Apr. 2011.

[60] A. Crabit, F. Colin, and R. Moussa, ''A soft hydrological monitoring approach for comparing runoff on a network of small poorly gauged catchments,'' *Hydrological Processes*, vol. 25, no. 18, pp. 2785–2800, Aug. 2011.

[61] X. C. Liu, T. C. Gao, and L. Liu, ''A comparison of rainfall measurements from multiple instruments,'' *Atmos. Meas. Techn.*, vol. 6, no. 7, pp. 1585–1595, Jul. 2013.

[62] L. Stevanato, G. Baroni, Y. Cohen, C. L. Fontana, S. Gatto, M. Lunardon, F. Marinello, S. Moretto, and L. Morselli, ''A novel cosmic-ray neutron sensor for soil moisture estimation over large areas,'' *Agriculture*, vol. 9, no. 9, p. 202, Sep. 2019.

[63] Z. Sokol, J. Szturc, J. Orellana-Alvear, J. Popová, A. Jurczyk, and R. Célleri, ''The role of weather radar in rainfall estimation and its application in meteorological and hydrological modelling—A review,'' *Remote Sens.*, vol. 13, no. 3, p. 351, Jan. 2021.

[64] H. Messer, A. Zinevich, and P. Alpert, ''Environmental monitoring by wireless communication networks,'' *Science*, vol. 312, no. 5774, p. 713, May 2006.

[65] H. Leijnse, R. Uijlenhoet, and J. N. M. Stricker, ''Rainfall measurement using radio links from cellular communication networks,'' *Water Resour. Res.*, vol. 43, no. 3, p. 321, Mar. 2007.

[66] M. Castillo-Effer, D. H. Quintela, W. Moreno, R. Jordan, and W. Westhoff, ''Wireless sensor networks for flash-flood alerting,'' in *Proc. 4th IEEE Int. Caracas Conf. Devices, Circuits Syst.*, Mar. 2004, pp. 142–146.

[67] Q. Abdelal and A. Al-Hmoud, ''Low-cost, low-energy, wireless hydrological monitoring platform: Design, deployment, and evaluation,'' *J. Sensors*, vol. 2021, pp. 1–14, Feb. 2021.

[68] Z. Zhang, S. D. Glaser, R. C. Bales, M. Conklin, R. Rice, and D. G. Marks, ''Technical report: The design and evaluation of a basin-scale wireless sensor network for mountain hydrology,'' *Water Resour. Res.*, vol. 53, no. 5, pp. 4487–4498, May 2017.

[69] D. A. Segovia-Cardozo, L. Rodríguez-Sinobas, F. Canales-Ide, and S. Zubelzu, ''Design and field implementation of a low-cost, open-hardware platform for hydrological monitoring,'' *Water*, vol. 13, no. 21, p. 3099, Nov. 2021.

[70] H. Vereecken, J. A. Huisman, Y. Pachepsky, C. Montzka, J. van der Kruk, H. Bogena, L. Weihermüller, M. Herbst, G. Martinez, and J. Vanderborght, ''On the spatio-temporal dynamics of soil moisture at the field scale,'' *J. Hydrol.*, vol. 516, pp. 76–96, Aug. 2014.

[71] H. Mittelbach and S. I. Seneviratne, ''A new perspective on the spatio-temporal variability of soil moisture: Temporal dynamics versus time-invariant contributions,'' *Hydrol. Earth Syst. Sci.*, vol. 16, no. 7, pp. 2169–2179, Jul. 2012.

[72] I. Schröter, H. Paasche, P. Dietrich, and U. Wollschläger, ''Estimation of catchment-scale soil moisture patterns based on terrain data and sparse TDR measurements using a fuzzy C-means clustering approach,'' *Vadose Zone J.*, vol. 14, no. 11, 2015, doi: 10.2136/vzj2015.01.0008.

[73] I. Wiekenkamp, J. A. Huisman, H. R. Bogena, H. S. Lin, and H. Vereecken, ''Spatial and temporal occurrence of preferential flow in a forested headwater catchment,'' *J. Hydrol.*, vol. 534, pp. 139–149, Mar. 2016.

[74] A. Salam, M. C. Vuran, and S. Irmak, ''Di-sense: In situ real-time permittivity estimation and soil moisture sensing using wireless underground communications,'' *Comput. Netw.*, vol. 151, pp. 31–41, Mar. 2019.

[75] A. Zanella, S. Zubelzu, M. Bennis, M. Capuzzo, and P. Tarolli, ''Internet of Things for hydrology: Potential and challenges,'' in *Proc. 18th Wireless On-Demand Netw. Syst. Services Conf. (WONS)*, Jan. 2023, pp. 114–121.

[76] D.-J. Seo, E. Habib, H. Andrieu, and E. Morin, ''Hydrologic applications of weather radar,'' *J. Hydrol.*, vol. 531, pp. 231–233, Dec. 2015.

[77] B. L. Barge, R. G. Humphries, S. J. Mah, and W. K. Kuhnke, ''Rainfall measurements by weather radar: Applications to hydrology,'' *Water Resour. Res.*, vol. 15, no. 6, pp. 1380–1386, Dec. 1979.

[78] A. Berne and W. F. Krajewski, ''Radar for hydrology: Unfulfilled promise or unrecognized potential?'' *Adv. Water Resour.*, vol. 51, pp. 357–366, Jan. 2013.

[79] E. Babaeian, M. Sadeghi, S. B. Jones, C. Montzka, H. Vereecken, and M. Tuller, ''Ground, proximal, and satellite remote sensing of soil moisture,'' *Rev. Geophys.*, vol. 57, pp. 530–616, Jun. 2019.

[80] F. Chen, Y. Gao, Y. Wang, and X. Li, ''A downscaling-merging method for high-resolution daily precipitation estimation,'' *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124414.

[81] J. Dong, L. Wei, X. Chen, Z. Duan, and Y. Lu, ''An instrument variable based algorithm for estimating cross-correlated hydrological remote sensing errors,'' *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124413.

[82] B. Mohammadi, ''Application of machine learning and remote sensing in hydrology,'' *Sustainability*, vol. 14, no. 13, p. 7586, Jun. 2022.

[83] H. Mosaffa, M. Sadeghi, I. Mallakpour, M. N. Jahromi, and H. R. Pourghasemi, ''Chapter 43—Application of machine learning algorithms in hydrology,'' in *Computers in Earth and Environmental Sciences*. Amsterdam, The Netherlands: Elsevier, 2022, pp. 585–591.

[84] B. Ghanbarian and Y. Pachepsky, ''Machine learning in vadose zone hydrology: A flashback,'' *Vadose Zone J.*, vol. 21, no. 4, Jul. 2022, Art. no. e20212.

[85] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, ''Ensemble machine learning paradigms in hydrology: A review,'' *J. Hydrol.*, vol. 598, Jul. 2021, Art. no. 126266.

[86] M. Ali, R. Prasad, Y. Xiang, and Z. M. Yaseen, ''Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts,'' *J. Hydrol.*, vol. 584, May 2020, Art. no. 124647.

[87] J. Farajzadeh and F. A. Alizadeh, ''A hybrid linear-nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model,'' *J. Appl. Meteorol.*, vol. 42, pp. 381–388, Jan. 2018.

[88] M. Ivković, A. Todorović, and J. Plavšić, ''Improved input to distributed hydrologic model in areas with sparse subdaily rainfall data using multivariate daily rainfall disaggregation,'' *J. Hydroinformatics*, vol. 20, no. 4, pp. 784–797, Jul. 2018.

[89] C. Mueller, T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee, ''NCAR auto-nowcast system,'' *Weather Forecasting*, vol. 18, no. 4, pp. 545–561, Aug. 2003.

[90] R. Rasmussen, M. Dixon, S. Vasiloff, F. Hage, S. Knight, J. Vivekanandan, and M. Xu, ''Snow nowcasting using a real-time correlation of radar reflectivity with snow gauge accumulation,'' *J. Appl. Meteorol.*, vol. 42, pp. 20–36, Jan. 2003.

[91] B. J. Turner, I. Zawadzki, and U. Germann, ''Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE),'' *J. Appl. Meteorol.*, vol. 43, no. 2, pp. 231–248, Feb. 2004.

[92] S. Ryu, G. Lyu, Y. Do, and G. Lee, ''Improved rainfall nowcasting using Burgers' equation,'' *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124140.

[93] N. I. Fox and C. K. Wikle, ''A Bayesian quantitative precipitation nowcast scheme,'' *Weather Forecasting*, vol. 20, no. 3, pp. 264–275, Jun. 2005.

[94] Y. Tikhamarine, D. Souag-Gamane, A. Najah Ahmed, O. Kisi, and A. El-Shafie, ''Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey wolf optimization (GWO) algorithm,'' *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124435.

[95] G. Zuo, J. Luo, N. Wang, Y. Lian, and X. He, ''Two-stage variational mode decomposition and support vector regression for streamflow forecasting,'' *Hydrol. Earth Syst. Sci.*, vol. 24, no. 11, pp. 5491–5518, Nov. 2020.

[96] A. Shabri, ''Streamflow forecasting using least-squares support vector machines,'' *Hydrological Sci. J.*, vol. 57, no. 7, pp. 1275–1293, Oct. 2012.

[97] Q.-T. Bui, Q.-H. Nguyen, X. L. Nguyen, V. D. Pham, H. D. Nguyen, and V.-M. Pham, ''Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping,'' *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124379.

[98] Y. Wang, Z. Fang, H. Hong, and L. Peng, ''Flood susceptibility mapping using convolutional neural network frameworks,'' *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124482.

[99] H. R. Pourghasemi, S. V. Razavi-Termeh, N. Kariminejad, H. Hong, and W. Chen, ''An assessment of metaheuristic approaches for flood assessment,'' *J. Hydrol.*, vol. 582, Mar. 2020, Art. no. 124536.

[100] A. Bhusal, U. Parajuli, S. Regmi, and A. Kalra, ''Application of machine learning and process-based models for rainfall-runoff simulation in DuPage river basin, Illinois,'' *Hydrology*, vol. 9, no. 7, p. 117, Jun. 2022.

[101] A. B. Dariane and S. Azimi, ''Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection,'' *J. Hydroinformatics*, vol. 20, no. 2, pp. 520–532, Mar. 2018.

[102] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, and S. Hochreiter, ''Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets,'' *Hydrol. Earth Syst. S.*, vol. 23, no. 12, pp. 589–611, 2019.

[103] M. J. Alizadeh, V. Nourani, M. Mousavimehr, and M. R. Kavianpour, ''Wavelet-IANN model for predicting flow discharge up to several days and months ahead,'' *J. Hydroinformatics*, vol. 20, no. 1, pp. 134–148, Jan. 2018.

[104] Y. Zhou, Z. Cui, K. Lin, S. Sheng, H. Chen, S. Guo, and C.-Y. Xu, ''Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques,'' *J. Hydrol.*, vol. 604, Jan. 2022, Art. no. 127255.

[105] U. Lall, ''Debates–The future of hydrological sciences: A (common) path forward? One water. One world. Many climes. Many souls,'' *Water Resour. Res.*, vol. 50, no. 6, pp. 5335–5341, Jun. 2014.

[106] M. Amjad, M. T. Yilmaz, I. Yucel, and K. K. Yilmaz, ''Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography,'' *J. Hydrol.*, vol. 584, May 2020, Art. no. 124707.

[107] L. C. Sieck, S. J. Burges, and M. Steiner, ''Challenges in obtaining reliable measurements of point rainfall,'' *Water Resour. Res.*, vol. 43, no. 1, p. 142, Jan. 2007.

[108] R. Hermann and A. Krener, ''Nonlinear controllability and observability,'' *IEEE Trans. Autom. Control*, vol. AC-22, no. 5, pp. 728–740, Oct. 1977.

[109] S. Omatu and T. Soeda, ''Optimal sensor location in a linear distributed parameter system,'' *IFAC Proc. Volumes*, vol. 10, no. 7, pp. 233–240, 1977.

[110] M. Basseville, A. Benveniste, G. Moustakides, and A. Rougee, ''Optimal sensor location for detecting changes in dynamical behavior,'' *IEEE Trans. Autom. Control*, vol. AC-32, no. 12, pp. 167–175, Dec. 1987.

[111] D. J. Chmielinski, T. Palmer, and V. Manousiouthakis, ''On the theory of optimal sensor placement,'' *AIChE J.*, vol. 48, no. 5, pp. 110–112, 2002.

[112] A. A. Alonso, I. G. Kevrekidis, J. R. Banga, and C. E. Frouzakis, ''Optimal sensor location and reduced order observer design for distributed process systems,'' *Comput. Chem. Eng.*, vol. 28, nos. 1–2, pp. 27–35, Jan. 2004.

[113] K. R. Muske and C. Georgakis, ''Optimal measurement system design for chemical processes,'' *AIChE J.*, vol. 49, no. 6, pp. 1488–1494, Jun. 2003.

[114] P. Sen, K. Sen, and U. M. Diwekar, ''A multi-objective optimization approach to optimal sensor location problem in IGCC power plants,'' *Appl. Energy*, vol. 181, pp. 527–539, Nov. 2016.

[115] U. Diwekar and R. Mukherjee, ''Optimizing spatiotemporal sensors placement for nutrient monitoring: A stochastic optimization framework,'' *Clean Technol. Environ. Policy*, vol. 19, no. 9, pp. 2305–2316, Nov. 2017.

[116] A. Afshar and M. A. Mariño, ''Multi-objective coverage-based ACO model for quality monitoring in large water networks,'' *Water Res. Manag.*, vol. 26, pp. 2159–2176, Jun. 2012.

[117] J. W. Berry, L. Fleischer, W. E. Hart, C. A. Phillips, and J. P. Watson, ''Sensor placement in municipal water networks,'' *J. Water Resour. Planning Manag.*, vol. 132, no. 3, pp. 192–203, 2005.

[118] R. Mukherjee, U. M. Diwekar, and A. Vaseashta, ''Optimal sensor placement with mitigation strategy for water network systems under uncertainty,'' *Comput. Chem. Eng.*, vol. 103, pp. 91–102, Aug. 2017.

[119] K. Worden and A. P. Burrows, ''Optimal sensor placement for fault detection,'' *Eng. Struct.*, vol. 26, pp. 885–901, Aug. 2001.

[120] W. H. Asquith, ''The use of support vectors from support vector machines for hydrometeorologic monitoring network analyses,'' *J. Hydrol.*, vol. 583, Apr. 2020, Art. no. 124522.

[121] S. Zubelzu, L. Rodríguez-Sinobas, D. Segovia-Cardozo, and A. Díez-Herrero, ''Optimal locations for flow and velocity sensors along a river channel,'' *Hydrological Sci. J.*, vol. 65, no. 5, pp. 800–812, Apr. 2020.

[122] M. D. Coles, D. Azzi, B. P. Haynes, and A. Hewitt, ''A Bayesian network approach to a biologically inspired motion strategy for mobile wireless sensor networks,'' *Ad Hoc Netw.*, vol. 7, no. 6, pp. 1217–1228, Aug. 2009.

[123] L. Miao, H. Qi, and F. Wang, ''Biologically-inspired self-deployable heterogeneous mobile sensor networks,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 2362–2368.

[124] G. Wang, G. Cao, and T. F. La Porta, ''Biologically-inspired self-deployable heterogeneous mobile sensor networks,'' *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 640–652, Aug. 2006.

[125] Z. Butler and D. Rus, ''Event-based motion control for mobile-sensor networks,'' *IEEE Pervasive Comput.*, vol. 2, no. 4, pp. 34–42, Oct. 2003.

[126] C. Pielli, D. Zucchetto, A. Zanella, and M. Zorzi, ''An interference-aware channel access strategy for WSNs exploiting temporal correlation,'' *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8585–8597, Dec. 2019.

[127] Z. Guo, M. Zhou, and G. Jiang, ''Adaptive sensor placement and boundary estimation for monitoring mass objects,'' *IEEE Trans. Syst., Man, Cybern., B*, vol. 38, no. 1, pp. 222–232, Feb. 2008.

[128] Y. Liu, Q. Zhao, W. Yao, X. Ma, Y. Yao, and L. Liu, ''Short-term rainfall forecast model based on the improved BP–NN algorithm,'' *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019.

[129] D. Sun, J. Wu, H. Huang, R. Wang, F. Liang, and H. Xinhua, ''Prediction of short-time rainfall based on deep learning,'' *Math. Problems Eng.*, vol. 2021, pp. 1–8, Mar. 2021.

[130] D. Pirone, L. Cimorelli, G. Del Giudice, and D. Pianese, ''Short-term rainfall forecasting using cumulative precipitation fields from station data: A probabilistic machine learning approach,'' *J. Hydrol.*, vol. 617, Feb. 2023, Art. no. 128949.

[131] S.-H. Moon, Y.-H. Kim, Y. H. Lee, and B.-R. Moon, ''Application of machine learning to an early warning system for very short-term heavy rainfall,'' *J. Hydrol.*, vol. 568, pp. 1042–1054, Jan. 2019.

[132] F. R. Adaryani, S. Jamshid Mousavi, and F. Jafari, ''Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN,'' *J. Hydrol.*, vol. 614, Nov. 2022, Art. no. 128463.

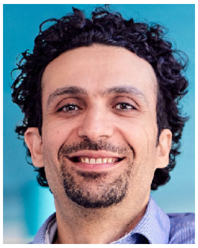
[133] V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N. V. Kalanker, M. R. Prakash, and R. Chand, ''Modeling and prediction of rainfall using artificial neural network and ARIMA techniques,'' *J. Ind. Geophys. Union*, vol. 10, no. 2, pp. 141–151, 2006.

[134] Y. Xu, C. Hu, Q. Wu, S. Jian, Z. Li, Y. Chen, G. Zhang, Z. Zhang, and S. Wang, ''Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation,'' *J. Hydrol.*, vol. 608, May 2022, Art. no. 127553.

[135] Y. Shao, J. Zhao, J. Xu, A. Fu, and M. Li, "Application of rainfall-runoff simulation based on the NARX dynamic neural network model," *Water*, vol. 14, no. 13, p. 2082, Jun. 2022.
[136] Z. Xiang, J. Yan, and I. Demir, "A rainfall-runoff model with LSTM-based sequence-to-sequence learning," *Water Resour. Res.*, vol. 56, no. 1, 2020, Art. no. e2019WR025326.
[137] S. P. Van, H. M. Le, D. V. Thanh, T. D. Dang, H. H. Loc, and D. T. Anh, "Deep learning convolutional neural network in rainfall-runoff modelling," *J. Hydroinform.*, vol. 22, no. 3, pp. 541–561, 2020.
[138] L. Schoppa, M. Disse, and S. Bachmair, "Evaluating the performance of random forest for large-scale flood discharge simulation," *J. Hydrol.*, vol. 590, Nov. 2020, Art. no. 125531.
[139] R. Zhang, J. Liu, H. Gao, and G. Mao, "Can multi-objective calibration of streamflow guarantee better hydrological model accuracy?" *J. Hydroinform.*, vol. 2, no. 3, pp. 687–698, 2018.
[140] M. Jay-Allemand, P. Javelle, I. Gejadze, P. Arnaud, P.-O. Malaterre, J.-A. Fine, and D. Organde, "On the potential of variational calibration for a fully distributed hydrological model: Application on a Mediterranean catchment," *Hydrol. Earth Syst. Sci.*, vol. 24, no. 11, pp. 5519–5538, Nov. 2020.
[141] X. Zhang and P. Liu, "A time-varying parameter estimation approach using split-sample calibration based on dynamic programming," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 2, pp. 711–733, Feb. 2021.
[142] J. Zaherpour, N. Mount, S. N. Gosling, R. Dankers, S. Eisner, D. Gerten, X. Liu, Y. Masaki, H. M. Schmied, Q. Tang, and Y. Wada, "Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models," *Environ. Modell. Softw.*, vol. 114, pp. 12–28, Apr. 2019.
[143] S. Grimaldi, A. Petroselli, F. Tauro, and M. Porfiri, "Time of concentration: A paradox in modern hydrology," *Hydrological Sci. J.*, vol. 57, no. 2, pp. 217–228, Feb. 2012.
[144] S. Grimaldi, A. Petroselli, and F. Nardi, "A parsimonious geomorphological unit hydrograph for rainfall–runoff modelling in small ungauged basins," *Hydrological Sci. J.*, vol. 57, no. 1, pp. 73–83, Jan. 2012.
[145] P. Jiang, P. Shuai, A. Sun, M. K. Mudunuru, and X. Chen, "Knowledge-informed deep learning for hydrological model calibration: An application to Coal Creek Watershed in Colorado," *Hydrol. Earth Syst. Sci.*, vol. 27, no. 14, pp. 2621–2644, 2022.
[146] M. K. Mudunuru, K. Son, P. Jiang, and X. Chen, "SWAT watershed model calibration using deep learning," 2021, *arXiv:2110.03097*.
[147] W.-P. Tsai, D. Feng, M. Pan, H. Beck, K. Lawson, Y. Yang, J. Liu, and C. Shen, "From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling," *Nature Commun.*, vol. 12, no. 1, p. 5988, Oct. 2021.
[148] Y. Fang, Y. Huang, B. Qu, X. Zhang, T. Zhang, and D. Xia, "Estimating the routing parameter of the Xin'anjiang hydrological model based on remote sensing data and machine learning," *Remote Sens.*, vol. 14, no. 18, p. 4609, Sep. 2022.
[149] J. Marçais and J.-R. de Dreuzy, "Prospective interest of deep learning for hydrological inference," *Groundwater*, vol. 55, no. 5, pp. 688–692, Sep. 2017.
[150] E. A. Varouchakis, D. T. Hristopulos, and G. P. Karatzas, "Improving Kriging of groundwater level data using nonlinear normalizing transformations—A field application," *Hydrological Sci. J.*, vol. 57, no. 7, pp. 1404–1419, Oct. 2012.
[151] H. Wu, Q. Yang, J. Liu, and G. Wang, "A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China," *J. Hydrol.*, vol. 584, May 2020, Art. no. 124664.
[152] K. Roushangar, F. Alizadeh, and V. Nourani, "Improving capability of conceptual modeling of watershed rainfall-runoff using hybrid wavelet-extreme learning machine approach," *J. Hydroinformatics*, vol. 20, no. 1, pp. 69–87, Jan. 2018.
[153] Y. Wang, M. Shao, Z. Liu, and D. N. Warrington, "Regional spatial pattern of deep soil water content and its influencing factors," *Hydrological Sci. J.*, vol. 57, no. 2, pp. 265–281, Feb. 2012.
[154] S. Das and C. Cunnane, "Performance of flood frequency pooling analysis in a low CV context," *Hydrological Sci. J.*, vol. 57, no. 3, pp. 433–444, Apr. 2012.
[155] J. D. Wolfe, K. R. Shook, C. Spence, and C. J. Whitfield, "A watershed classification approach that looks beyond hydrology: Application to a semi-arid, agricultural region in Canada," *Hydrol. Earth Syst. Sci.*, vol. 23, no. 9, pp. 3945–3967, Sep. 2019.

ANDREA ZANELLA (Senior Member, IEEE) is a Full Professor with the Department of Information Engineering (DEI), University of Padova, Italy. His long-established research activities are in protocol design, optimization, and performance evaluation of wired and wireless networks, with 200+ publications in top journals and conferences in the sector. He was the Technical Area Editor of the IEEE Internet of Things Journal, and an Associate Editor of the IEEE Transactions on Cognitive Communications and Networking, IEEE Communications Surveys and Tutorials, and *Digital Communications and Networks* (DCN).

SERGIO ZUBELZU is a Professor with the Department of Agroforest Engineering, Technical University of Madrid. Since 2021, he has been the Deputy Director of Infrastructures with the School of Agriculture, Food and Biosystems Engineering. With more than 20 years of experience in both private sector and academia, he has published more than 20 articles in highly ranked scientific journals and led more than 15 research projects. His research interests span both hydrological monitoring and modeling with focus on coupling physical and data-driven approaches.



MEHDI BENNIS (Fellow, IEEE) is a full-tenured Professor with the Centre for Wireless Communications, University of Oulu, Finland, and the Head of the Intelligent Connectivity and Networks/Systems Group (ICON). His main research interests are in radio resource management, game theory, and distributed AI in 5G/6G networks. He has published more than 200 research papers in international conferences, journals, and book chapters. He was a recipient of several prestigious awards. He is an Editor of IEEE Transactions on Communications and a Specialty Chief Editor for Data Science for Communications of the *Frontiers in Communications and Networks* journal.