

Can Dominant Runoff Generation Mechanisms Be Disentangled Through Hypothesis Testing? Insights From Integrated Hydrological-Hydrodynamic Modeling

Perrini, Pasquale; Iacobellis, Vito; Gioia, Andrea; Cea, Luis; Savenije, Hubert H.G.; Fenicia, Fabrizio

DOI

[10.1029/2024WR039394](https://doi.org/10.1029/2024WR039394)

Publication date

2025

Document Version

Final published version

Published in

Water Resources Research

Citation (APA)

Perrini, P., Iacobellis, V., Gioia, A., Cea, L., Savenije, H. H. G., & Fenicia, F. (2025). Can Dominant Runoff Generation Mechanisms Be Disentangled Through Hypothesis Testing? Insights From Integrated Hydrological-Hydrodynamic Modeling. *Water Resources Research*, 61(4), Article e2024WR039394. <https://doi.org/10.1029/2024WR039394>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Water Resources Research®



RESEARCH ARTICLE

10.1029/2024WR039394

Key Points:

- Surface and subsurface runoff mechanisms can alternatively dominate different flood events in semi-arid regions
- A landscape-based model structure provides enhanced performance and realistic simulated internal processes
- Hypothesis testing and perceptual insights can compensate for limited data when only streamflow is available

Correspondence to:

P. Perrini,
pasquale.perrini@uniba.it

Citation:

Perrini, P., Iacobellis, V., Gioia, A., Cea, L., Savenije, H. H. G., & Fenicia, F. (2025). Can dominant runoff generation mechanisms be disentangled through hypothesis testing? insights from integrated hydrological-hydrodynamic modeling. *Water Resources Research*, 61, e2024WR039394. <https://doi.org/10.1029/2024WR039394>

Received 14 NOV 2024

Accepted 7 APR 2025

Author Contributions:

Conceptualization: Pasquale Perrini, Vito Iacobellis, Hubert H. G. Savenije, Fabrizio Fenicia

Data curation: Pasquale Perrini

Formal analysis: Pasquale Perrini

Investigation: Pasquale Perrini

Methodology: Pasquale Perrini,

Vito Iacobellis, Luis Cea, Hubert

H. G. Savenije, Fabrizio Fenicia

Software: Pasquale Perrini, Luis Cea

Supervision: Vito Iacobellis,

Andrea Gioia, Luis Cea, Fabrizio Fenicia

Writing – original draft: Pasquale Perrini

Writing – review & editing:

Vito Iacobellis, Andrea Gioia, Luis Cea,

Hubert H. G. Savenije, Fabrizio Fenicia

Can Dominant Runoff Generation Mechanisms Be Disentangled Through Hypothesis Testing? Insights From Integrated Hydrological-Hydrodynamic Modeling

Pasquale Perrini^{1,2,3} , Vito Iacobellis², Andrea Gioia², Luis Cea⁴ , Hubert H. G. Savenije⁵, and Fabrizio Fenicia³ 

¹Department of Soil, Plant and Food Science, University of Bari Aldo Moro, Bari, Italy, ²Department of Civil, Environmental, Land, Building Engineering and Chemistry, Polytechnic University of Bari, Bari, Italy, ³Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, ⁴Water and Environmental Engineering Group, Center for Technological Innovation in Construction and Civil Engineering (CITEEC), University of A Coruña, A Coruña, Spain, ⁵Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, CN Delft, The Netherlands

Abstract Identifying flood-inducing processes remains a challenge in catchment hydrology due to the complex runoff dynamics, particularly in semi-arid regions where surface and subsurface mechanisms alternatively drive streamflow across seasons. Tracer data can help identify hydrograph sources, but they are often unavailable or lack sufficient temporal resolution. To aid process identification at the event-scale, we developed an integrated hydrological-hydrodynamic framework and compared multiple model hypotheses informed by hydrological signatures. We systematically tested these hypotheses through falsification, meta-evaluation, spatial validation, and posterior diagnostics, using the semi-arid Salsola nested catchment in southern Italy as case study. While all model structures performed well on common calibration metrics, differences emerged in spatial transferability tests and alternative diagnostic assessments. Some models, despite strong performance, exhibited inconsistent representations of internal runoff mechanisms, indicating that they achieved good results for the wrong reasons. Furthermore, the choice of routing schemes significantly influenced high-peak estimations and overall model performance, particularly when Horton-type overland flow was considered. This underscores the need to treat routing methods as a key component in event-scale modeling. Our findings reveal that during consecutive storm events in the study catchment, surface processes dominate the initial stages, whereas subsurface processes become more influential in later events, providing valuable insights that may be applicable to similar semi-arid regions. Overall, we emphasize the importance of hypothesis testing in runoff process identification, which can compensate for the absence of hydrochemical data for hydrograph separation. Additionally, our results highlight the value of a landscape-based modeling approach for distinguishing alternative runoff generation processes.

Plain Language Summary Predicting how runoff generates, sustains rivers, and triggers floods is challenging. In this study, we investigate the processes driving runoff in a semi-arid catchment and explore how they can be identified with limited data. Rather than relying on a fixed model, we tested multiple model structures, each representing different hypotheses about runoff generation. By evaluating how well these hypotheses transfer to similar catchments, we identified the most effective ones based on both predictive performance and physical consistency. Our results show that incorporating landscape heterogeneity improves model predictions and enhances streamflow estimation. This approach provides valuable insights into flood-inducing processes, particularly in data-scarce regions.

1. Introduction

Streamflow generation processes exhibit significant spatial and temporal variability across different landscapes and storm events, making them challenging to identify and predict (Kirchner, 2024; McDonnell et al., 2007). In humid catchments, hydrological responses are typically driven by catchment wetness, with processes such as saturation-excess overland flow and storage-excess subsurface flow playing key roles (Dunne & Black, 1970; Sayama et al., 2011). Conversely, in arid catchments, infiltration-excess overland flow is more prevalent, with streamflow responses primarily controlled by rainfall intensity (Horton, 1933; Tao et al., 2021). Pioneering studies have developed distinct models for arid and humid regions, explicitly conceptualizing differences in

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

runoff mechanisms (Zhao, 1977). More recently, multiple mechanisms and highly specific processes (e.g., fill-and-spill subsurface behavior) have been modeled at the catchment scale using integrated surface-subsurface hydrological models (ISSHM) (Paniconi & Putti, 2015), namely the latest generation of physically based models (Camporese et al., 2019; Maxwell et al., 2014; Zanetti et al., 2024). Complementary data to the streamflow time series, such as stable isotopes or geochemical tracer concentrations, provide independent insights and allow cross-validation of modeled runoff contributions to flow hydrographs (Birkel et al., 2014; McMillan, Tetzlaff, et al., 2012). However, such data are rarely available, and often lack sufficient temporal and spatial resolution.

Characterizing and modeling catchment responses becomes increasingly difficult as one moves from coarser to finer time scales (Atkinson et al., 2002) and from wetter to drier catchments (Wheater et al., 2007). As temporal resolution increases, multiple surface runoff mechanisms become more apparent (McMillan, 2020), requiring finer sampling to detect short-duration processes (Kirchner et al., 2004). Concurrently, transitioning from long-term to event-scale modeling requires more complex models to accurately capture fast runoff dynamics, such as double peaked responses (Kavetski et al., 2011). Furthermore, detailed routing methods are necessary to improve streamflow representation, particularly at sub-daily scales (Cortés-Salazar et al., 2023). Unlike humid areas, where a large part of the hydrological models can accurately simulate flood processes (Sittner, 1976), event-scale modeling in arid and semi-arid regions is considerably more challenging (Al-Qurashi et al., 2008; Huang et al., 2016). This challenge arises from the high variability of hydrological regimes in semi-arid catchments, which fluctuate between dry and wet states, along with seasonal changes in catchment properties. For instance, soils may develop water repellency during the dry season, altering infiltration behavior (Doerr et al., 2000). This variability results in seasonal shifts in overland flows types, complicating the commonly assumed relationship between reduced infiltrability and increasing soil saturation.

From a modeling perspective, it is not always clear which specific runoff processes should be represented in semi-arid regions. In some cases, saturation-excess and subsurface processes are considered negligible (Kidron, 2021), while in others, they are thought to play a dominant role (Mul et al., 2008). When sufficient data exist for detailed physical parametrization, combining isotope analysis with ISSHM provides a robust approach for disentangling actual runoff generation mechanisms at the event-scale (Chen et al., 2023). However, such analyses are typically limited to well-instrumented headwater catchments, as high frequency hydrochemical data remains scarce, particularly in ephemeral stream networks, common to (semi) arid environments. It remains unclear whether streamflow data alone are sufficient to fully distinguish between alternating runoff generation processes.

In the absence of detailed prior knowledge of hydrological processes in (semi) arid regions, hypothesis testing through modeling becomes an essential, if not the sole, tool for inferring dominant processes at the mesoscale or regional scale (Astagneau et al., 2022; Beven, 2018; Fenicia et al., 2022). The paradigm of flexible modeling is particularly useful for exploring multiple testable hypotheses and retrospectively identifying the most likely catchment behavior (Clark et al., 2011). This can be achieved by combining different model structures, representing a set of decisions and assumptions regarding catchment discretization, hydrological processes, constitutive functions, mathematical methods, and free parameters used for calibration (Fenicia et al., 2016). Notable flexible modeling frameworks include SUPERFLEX (Dal Molin et al., 2020; Fenicia et al., 2011), FUSE (Clark et al., 2008), SUMMA (Clark et al., 2015) and MARRMoT (Knoben et al., 2019). These frameworks share the ability to generate alternative model structures but differ in their scope of application. For example, SUPERFLEX, MARRMoT and FUSE are more oriented toward conceptual models, while SUMMA is designed for physically based modeling.

Since runoff generation processes vary spatially and temporally, a distributed approach is essential for testing alternative landscape controls on runoff generation. Surface topography is a key factor influencing runoff, as many models incorporate it to explain spatial heterogeneity in hydrological responses. Conceptual frameworks and similarity indices that often link geomorphic properties to hydrological behavior/functioning (Goudarzi et al., 2023; Loritz et al., 2019; Nobre et al., 2011). However, multiple landscape classification approaches exist, each defining distinct runoff types (Antonetti et al., 2016).

One of the most widely used theories is the topographic index approach, forming the foundation of TOPMODEL and its evolutions (Beven et al., 2020). This theory posits that a location's contribution to streamflow increases with catchment wetness and contributing area, while decreasing with local slope. Alternatively, Savenije (2010) proposed that topographically similar areas within a catchment — such as wetlands, hillslopes, and plateaus —

exhibit distinct hydrological behaviors, forming the basis of the FLEX-topo model, which distinguishes between various landscape units and identifies unique dominant processes associated with each (Gao et al., 2016; Gharari et al., 2014).

Model evaluation plays a crucial role in identifying hydrological processes and ranking competing hypotheses (Fenicia & Kavetski, 2021). Uncertainties in traditional quantitative metrics necessitate the use of system-scale metrics (meta-metrics) that promote hydrologically consistent parameter sets rather than purely statistical accuracy (Clark et al., 2021; Zhang et al., 2008). While split-sample calibration/validation remains a widely used hydrological modeling practice (Klemeš, 1986) it may be insufficient for rigorously testing multiple working hypotheses (Shen et al., 2022). This highlights the need for more robust evaluation strategies, such as validating model structures based on spatial transferability (Gao et al., 2016; Gupta et al., 2014).

These considerations provide the premise for this study, which investigates the event-scale runoff generation processes in a semi-arid region at sub-hourly temporal resolution within a fully distributed modeling framework. The Salsola nested catchment in southern Italy serves as a case study, utilizing high-frequency rainfall-radar and streamflow data from four consecutive storm events, each yielding distinct hydrograph responses in time and space. We present an integrated modeling framework that combines distributed model structures based on the DREAM hydrological model (Perrini et al., 2024) for local runoff estimation with the Iber+ shallow water model for flow routing (Bladé et al., 2014; García-Feal et al., 2018). Unlike other flexible frameworks, our approach integrates directly a 2D hydrodynamic model, allowing joint evaluation of various distributed hydrological hypotheses and different runoff routing methods.

The main objective of this research is to disentangle surface and subsurface runoff contributions across different sub-catchments and consecutive storm events while developing a realistic model structure for the region. In the absence of isotope or geochemical tracer data, we employ multiple working hypotheses to determine the extent to which streamflow data alone can distinguish alternative runoff generation mechanisms. Furthermore, we illustrate how the relative merits of competing models depend on rigorous evaluation, which cannot be achieved solely through conventional quantitative metrics. Four model architectures, informed by hydrological signatures and prior catchment understanding, were tested. These structures represent different runoff generation assumptions, ranging from exclusive surface or subsurface processes to a combination of both, with further refinements based on landscape characteristics, consistent with the FLEX-topo framework (Savenije, 2010). We evaluate these hypotheses through a multi-stage performance analysis, employing diverse metrics to assess the internal consistency of the modeled processes.

1.1. Runoff Processes Taxonomy and Terminology

Motivated by the need for a unified taxonomy of hydrological processes (McMillan, 2022) and to prevent confusion within the scientific community, we explicitly define the three main runoff generation processes analyzed in this study. These processes are: Infiltration-excess overland flow (Horton-type overland flow, HOF), Saturation-excess overland flow (Dunne-type overland flow, SOF), and Storage-excess subsurface flow (McDonnell-type, SSF). The first two are classified as surface runoff processes, while the third is a subsurface process. All play a significant role in flood generation, though surface runoff processes (HOF and SOF) typically respond more rapidly than subsurface flow (SSF).

HOF occurs when rainfall intensity exceeds the soil infiltration capacity, leading to surface overland flow (Horton, 1933). This process does not require prior soil saturation and is most common in areas with low-permeability soils or intense rainfall events.

SOF, as conceptualized by Thomas Dunne (Dunne & Black, 1970), occurs in areas where the soil is already saturated, often near streams where a rising groundwater table reduces infiltration capacity. This process aligns with the Variable Contributing Area (VCA) concept, which describes the dynamic expansion and contraction of saturated zones that contribute to streamflow.

SSF is driven by subsurface dynamics under partially or fully saturated soil conditions (Weiler et al., 2006). It consists of lateral preferential flow paths and/or macropore flow, often exhibiting fill-and-spill behavior (McDonnell et al., 2021), where water accumulates in bedrock depressions and moves downslope once hydrological connectivity is established. Hereinafter, we refer to SSF only as the net amount of water exfiltrating from the subsurface and emerging at the land surface as a source of runoff.

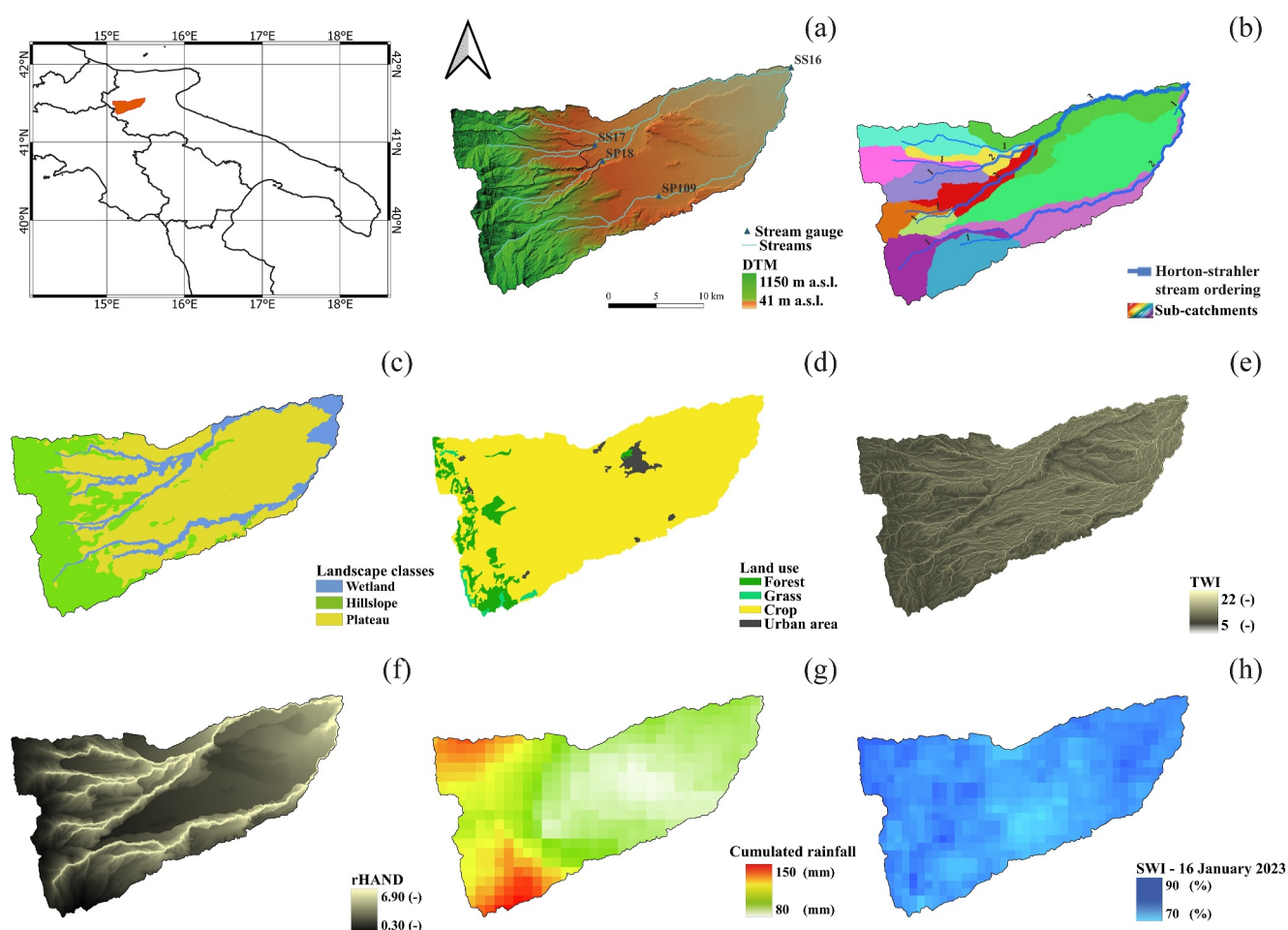


Figure 1. Salsola study case: (a) Digital Terrain Model (DTM) with elevation in meters above sea level (m a.s.l.), streamflow gauge locations and corresponding drainage area, (b) Horton-Strahler stream ordering and corresponding sub-catchments, (c) HAND-based landscape classification, (d) main land use classes, (e) Topographic wetness index (TWI), (f) rHAND index, (g) cumulated rainfall between the 16th and the 30 January 2023, (h) Soil Water Index (SWI) on the 16 January 2023.

2. Study Area

This research was conducted in the Salsola catchment (Figure 1), a tributary of the Candelaro River one of the largest rivers in the Apulia region located in southeastern Italy. The Salsola catchment covers approximately 430 km² until the section where the surrounding land drops below the river level (as a result of land subsidence) and no longer drains naturally to the main channel. Only 4% of the area is urbanized, with the primary land use being extensive agriculture in the upper part of the basin and intensive agriculture in the lowlands (Figure 1d). The top-soil composition varies significantly, ranging from sandy-clay-loam to clay-loam or clay. Whereas the average elevation is about 300 m above sea level, ranging from 41 to 1,150 m (Figure 1a).

The catchment includes three gauged headwater tributaries that are particularly flash-prone. One of these, the SP109 gauge station, has discontinuous measurements preventing the use of its data in this case study. The streamflow regime changes rapidly and closely follows the unevenly distributed precipitation. Orographic factors significantly influence both the amount and pattern of rainfall, often resulting in short high-intensity storm events. Overall, the Salsola exhibits typical Mediterranean semi-arid characteristics, with seasonal patterns of droughts and flash floods. Like other watercourses in the Candelaro system, the Salsola is classified as a temporary river, experiencing long periods of little to no flow (De Girolamo et al., 2015). The average streamflow at the Salsola outlet (i.e., SS16 gauge station) in recent years (2022–2023) is approximately 0.30 mm/day, while in the headwater catchments 0.80 mm/day at the SP18 station and 0.74 mm/day at the SS17 station.

Table 1

Long-Term and Event-Scale Hydrological Signatures: I_B (the Ratio of Baseflow Over Total Streamflow), I_F (the Ratio Between the Sum of Absolute Changes in Streamflow Over Consecutive Time Steps and the Total Streamflow), I_D (the Ratio Between Average Potential Evaporation and Average Precipitation), I_E (the Ratio Between Average Actual Evaporation, Calculated as the Difference Between Average Precipitation and Streamflow, and Average Precipitation), I_C (the Ratio Between Cumulated Rainfall and Streamflow)

Gauge station	Drainage area (km ²)	Long term					Event scale	
		I_B	I_F	I_D	I_E	I_C	I_F	I_C
SS16	431	0.11	0.53	1.65	0.59	0.17	0.05	0.38
SS17	55	0.10	0.51	1.40	0.63	0.36	0.10	0.45
SP18	42	0.11	0.32	1.46	0.62	0.40	0.13	0.50

For the purpose and context of this study, which aims to identify the dominant sources of runoff generation at the event-scale, we selected a 15 days period of streamflow data from 16 to 30 January, 2023. During this time, four distinct hydrographs were recorded at the Salsola SS16 outlet due to a series of consecutive storm events of different intensity along the days that provoked multiple overflowing in the final reach of the mainstream. Further details of this choice are explained in Section 4.1.

2.1. Data

All the hydrometric and meteorological data used for the construction of the data layers were provided by the Civil Protection of the Puglia Region, while the Rainfall-radar products are the one of the National departments of the Civil Protection. The distributed rainfall field (Figure 1g) aggregated each 15 min was estimated merging rain-gauge measurement with the radar layers exploiting the Kriging with external drift (Jewell & Gaussiat, 2015; Ochoa-Rodriguez et al., 2019). The reference topography is the 1 m resolution Lidar-based digital terrain model provided by the Italian Ministry of the Environment and Energy Security—Mase (available online at <https://sim.mase.gov.it/>). The grouped land uses classes were identified starting from the Corine Land Cover 2018 map, and corresponding manning coefficients were assigned following Chow (1959). The distributed taxonomy, depth and hydraulic parametrization of the soil is the one of the ACLA2 Apulian regional project (Caliandro et al., 2005), while the organic content is based on the European LUCAS database (de Brogniez et al., 2015). Potential evaporation has been estimated following the FAO procedure (Allen et al., 1998) interpolating the punctual measurement of wind speed and temperature with the ordinary kriging procedure. Vegetation dynamic over time has been considered varying the Leaf area index (LAI) every 4 days, namely the temporal resolution of the reference LAI-MODIS gridded products with 0.5 km spatial resolution (available online at <https://lpdaac.usgs.gov/products/mcd15a3hv061/>). Finally, the satellite detected soil moisture contents (Figure 1h), mentioned in Section 4.1, is the Soil Water Index T20 (I_{SWI} ; Bauer-Marschallinger et al., 2018) with a daily 1 km resolution which represent the normalized soil moisture content between the wilting point and the field capacity (available online at <https://land.copernicus.eu/en/products/soil-moisture>).

2.2. Hydrological Signatures and Prior Model Hypotheses

The value of incorporating certain (sub-)processes into models relies on their significance within the study area (Guse et al., 2021). To gain a deeper understanding of the spatial and temporal variability of streamflow responses in the Salsola catchments, and their potential causes, we analyzed streamflow signatures in the area. This analysis is also valuable during model development, as it helps inform the formulation of competing model hypotheses, and during model evaluation, providing a foundation for designing effective diagnostic tests (Fenicia & McDonnell, 2022).

Since some climate signatures have relevance over hydrologically significant timescales (years), we first assessed the signatures for the years 2022 and 2023 using daily averaged streamflow measurements. Specifically, we considered the Baseflow Index I_B , the Flashiness index I_F , the dryness index I_D , the evaporation index I_E , and the runoff coefficients I_C , computed as in Fenicia and McDonnell (2022) (Table 1).

In terms of long-term averages, I_F increases when moving from the headwaters to the outlet, indicating higher responsiveness to rainfall events in larger catchment areas. I_B is similar for all catchments, showing a baseflow

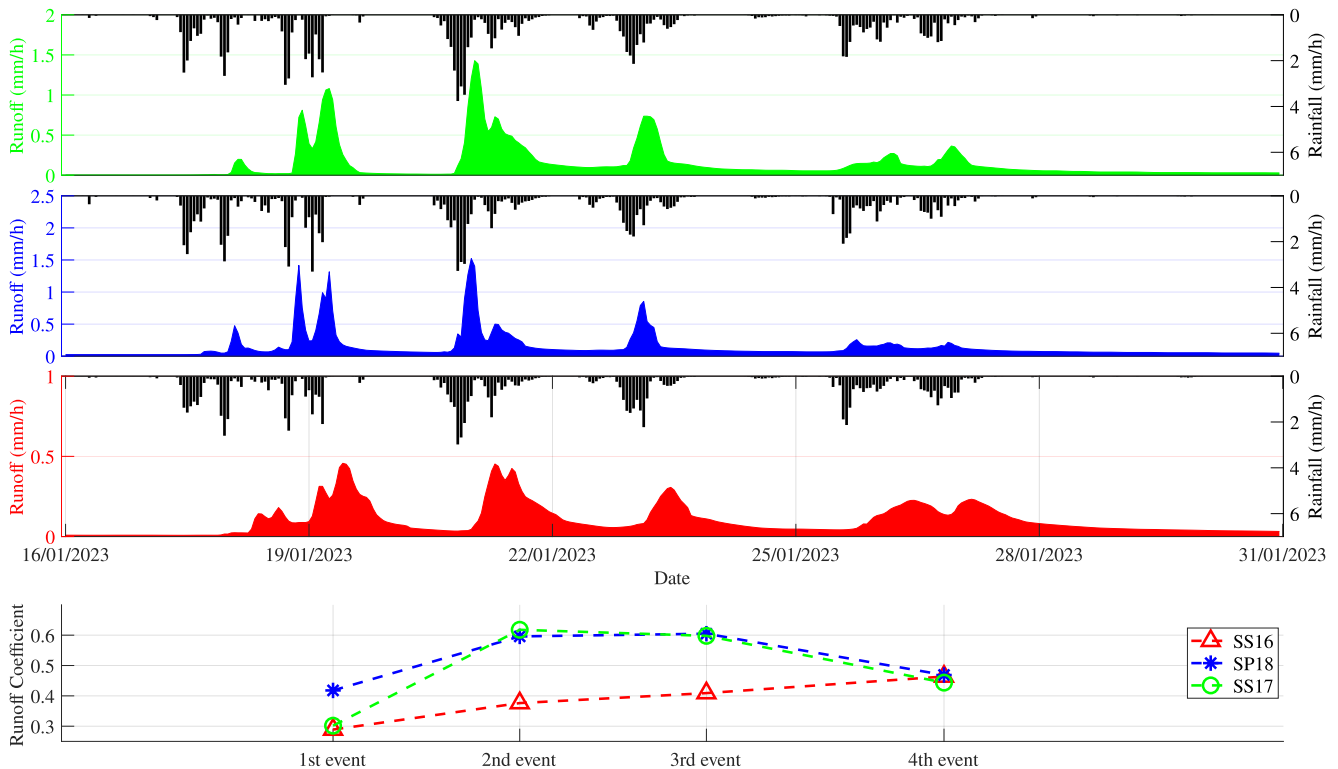


Figure 2. Rainfall and streamflow at each streamflow gauge station. The bottom panel shows the runoff coefficients for each storm event.

contribution around 10%, which is considered relatively low (Smakhtin, 2001). I_C decrease with increasing drainage area. I_D is significantly greater than one in all catchments while I_E remains around 0.60, indicating substantial evaporative fluxes with respect to incoming precipitation. Typical values of an arid or semi-arid region, as previously mentioned.

At the event scale only the I_F and I_C were calculated, as the other signatures are not particularly meaningful over such a short time span. Here, I_F decreases with catchment area, showing a reverse trend compared to the long-term behavior. I_C remains inversely dependent on the drainage area and is significantly higher than the long-term. To establish the existence or importance of baseflow at the event-scale, we conducted a visual inspection of the hydrographs as suggested by Wrede et al. (2015). For the two-headwater catchment, it is visible that the baseflow contribution is very limited, especially at the beginning of the period meaning that quick flow is dominant. The outlet instead, appears to have a more pronounced baseflow contribution, albeit still relatively low.

The event-scale runoff coefficients show an irregular pattern, whereas the outlet runoff coefficients display a continuous increase (Figure 2, bottom panel). Our prior interpretation is assuming reasonably that catchment wetness increases progressively, given that the four events occur close in time, and evaporation is minimal during the rainy winter season.

However, in the headwater catchments, infiltration-excess processes are potentially occurring since the temporal rainfall-runoff relations are related more to rainfall intensity than to catchment storage. While for the SS16 outlet we observed an always increasing rainfall-runoff relationship which suggests that other processes like SOF or SSF can be concurrent to the HOF (McMillan, 2020). These assumptions are also supported by a previous flood frequency analysis conducted in the area, which identified HOF as the dominant runoff mechanism during severe rainfall events (Iacobellis et al., 2011).

In summary, the analysis based on streamflow signatures and visual inspection described above suggests the following hypotheses regarding runoff-generating mechanisms for the events considered.

- The baseflow process is negligible compared to the quickflow contribution at the event-scale;
- The headwater catchments are characterized by the presence of HOF;

- The streamflow at the outlet is driven by multiple types of runoff;

These hypotheses will guide model decisions and inform the model comparison exercise, as outlined in Section 3.2.

3. Methodology

3.1. Hydrological-Hydrodynamic Framework

The modeling framework in this study integrates two components: a conceptual distributed hydrological model to identify runoff sources, and a physically based two-dimensional hydrodynamic model for routing. The hydrological model is spatially distributed, providing grid-based inputs for the hydrodynamic model. This allows it to capture the spatial variability of intense rainfall, which can lead to localized responses that lumped models cannot detect.

The use of a conceptual model keeps the process description simple, focusing on dominant processes. Specifically, the grid-based model simulates soil water budget using a bucket-type approach, often referred to as a distributed “integral” model (Clark et al., 2015). This approach uses a spatial assemblage of one-dimensional column models connected by lateral fluxes between individual columns.

Since it has been both theoretically and empirically shown that catchments responses exhibit significant variability to rainfall intensity (Grimaldi et al., 2012; Michailidi et al., 2018), instead of a hydrological routing scheme, we employ a hydrodynamic code based on the 2D depth-averaged Shallow Water Equations (2D-SWE). Indeed, the hydrodynamics is crucial for event-based applications and short time scales, since simplified routing methods struggle to predict the highly transient flooding caused by intense rainfall (Ming et al., 2020). Conversely, physically based hydraulic routing improves model realism, reducing structural uncertainties related to surface wave propagation. The considerations align with a growing branch of the hydrological community that utilizes catchment-scale models based on the 2D-SWE for event-based hindcasting or nowcasting (e.g., Cea et al., 2024; Costabile et al., 2023) and for estimating travel times with a more physical basis (e.g., Barbero et al., 2022; Garzon et al., 2023).

The overall framework proposed herein aims to balance realism in representing processes with the need to keep model complexity manageable, particularly in terms of the number of parameters involved, as discussed later in detail. In order to couple a grid-based hydrological model with a 2D-SWE model we adopted the recent developed Runoff-on-Grid approach which embed the amount of the runoff generation processes computed by the hydrological model as a spatiotemporal boundary condition of the 2D-SWE model, to provide a fully integrated hydrological-hydrodynamic framework at the catchment scale (Perrini et al., 2024). Here, reference 2D-SWE model is the open source GPU-enhanced software Iber+ (Bladé et al., 2014; García-Feal et al., 2018), which uses an efficient and stable finite volume solver, the DHD (Cea & Bladé, 2015), for catchment-scale hydrological applications. While the hydrological model is a modular framework which was built based on the MATLAB code of the first version of the DREAM model proposed by Manfreda et al. (2005).

All the hydrological model structures, presented in the next section, use the same spatial discretization fixed to a 50 m uniform grid and a time-step of 15 min, as well as the same computational set-up, mesh geometry and roughness parameters of the 2D-SWE model. In particular, the Iber+ hydrodynamic model uses an unstructured non-uniform discretization differentiated between the main streams and external areas. To improve the accuracy of the hydrodynamic results without increasing too much the total number of computational elements, we carried out a preliminary sensitivity analysis of different mesh resolutions. A 5 m resolution for the main streams and tributaries and 50 m for the external areas was adopted leading to a total of 800'000 computational elements. Actually, in rural areas very little performance are gained by running simulations at resolutions finer than 50 m, incurring unnecessary computational costs (Savage et al., 2016).

3.2. Model Structures

Four model structures of different complexities, that is, HORTONIAN, DUNNIAN, DREAM and FLEX-DREAM, have been selected to represent multiple testable hypotheses for the different runoff generation processes. It is worth noting that the term “model structure” refers to the hydrological model set-up, which feeds the 2D-hydrodynamic model that remains untouched.

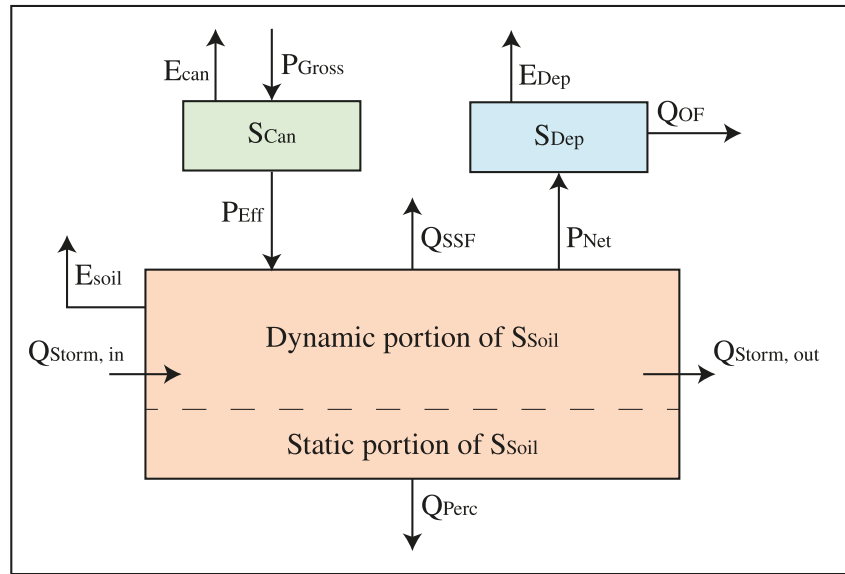


Figure 3. Generalized illustration of the ODEs.

To test solely the different hypotheses of runoff production, all bucket-type structures share the same climatic input, soil-vegetation-land use parametrizations, and the same modules for interception, surface depression, and actual evaporation. To reduce the number of calibration parameters while maintaining the spatial distribution of the data, we used regularization functions that link local parameters to global hyperparameters, which is a common procedure in distributed modeling applications (Fenicia et al., 2016; Pokhrel et al., 2008; Pokhrel & Gupta, 2010). In this specific case, the hyperparameters act as linear scaling factors of properties directly taken from the maps, while local parameters regulate the physical meaning of processes.

To simplify model description, all structures can be thought as being described by the same mass balance equations, differing in the presence or absence of specific fluxes. Hereafter, all the storages are identified with S [L], while all the fluxes with Q, P and E [LT⁻¹]. The mass balance of each landscape element (or pixel) of the distributed model is described by three ordinary differential equations (ODEs) solved sequentially with an explicit bounded Euler method. The state variables are S_{Can} canopy storage, S_{Dep} surface depression storage and S_{Soil} soil storage. The ODEs are expressed as follows:

$$\frac{dS_{Can}}{dt} = P_{Gross} - P_{Eff} - E_{Can} \quad (1)$$

$$\frac{dS_{Soil}}{dt} = P_{Eff} - P_{Net} - Q_{Perc} + Q_{Storm,in} - Q_{Storm,out} - Q_{SSF} - E_{Soil} \quad (2)$$

$$\frac{dS_{Dep}}{dt} = P_{Net} - Q_{OF} - E_{Dep} \quad (3)$$

where P_{Gross} is the gross rainfall, P_{Eff} is the effective rainfall (gross rainfall minus canopy interception), E_{Can} is the direct evaporation from canopy, P_{Net} is the net rainfall-excess, Q_{Perc} is the percolation out of the soil column, $Q_{Storm,in}$ and $Q_{Storm,out}$ are the subsurface fluxes entering and leaving the single grid cell, Q_{SSF} is the storage-excess subsurface flow, E_{Soil} is the actual evaporation from the soil, Q_{OF} is the direct overland flow (i.e., either HOF or SOF depending on the saturation degree of the single cell) and finally E_{Dep} is the direct evaporation from surface depression. The soil water content at saturation ($S_{Soil,max}$), at field capacity ($S_{Soil,c}$), at wilting point ($S_{Soil,wp}$) and the residual ($S_{Soil,r}$) are computed as the product of the corresponding volumetric soil moisture content (θ [-]) and the soil depth (D [L]).

In the next subsections, each model architecture is described in order of complexity, with emphasis on the singular peculiarities and assumptions. Figure 3 provides a conceptual illustration of the buckets and fluxes, aiding the

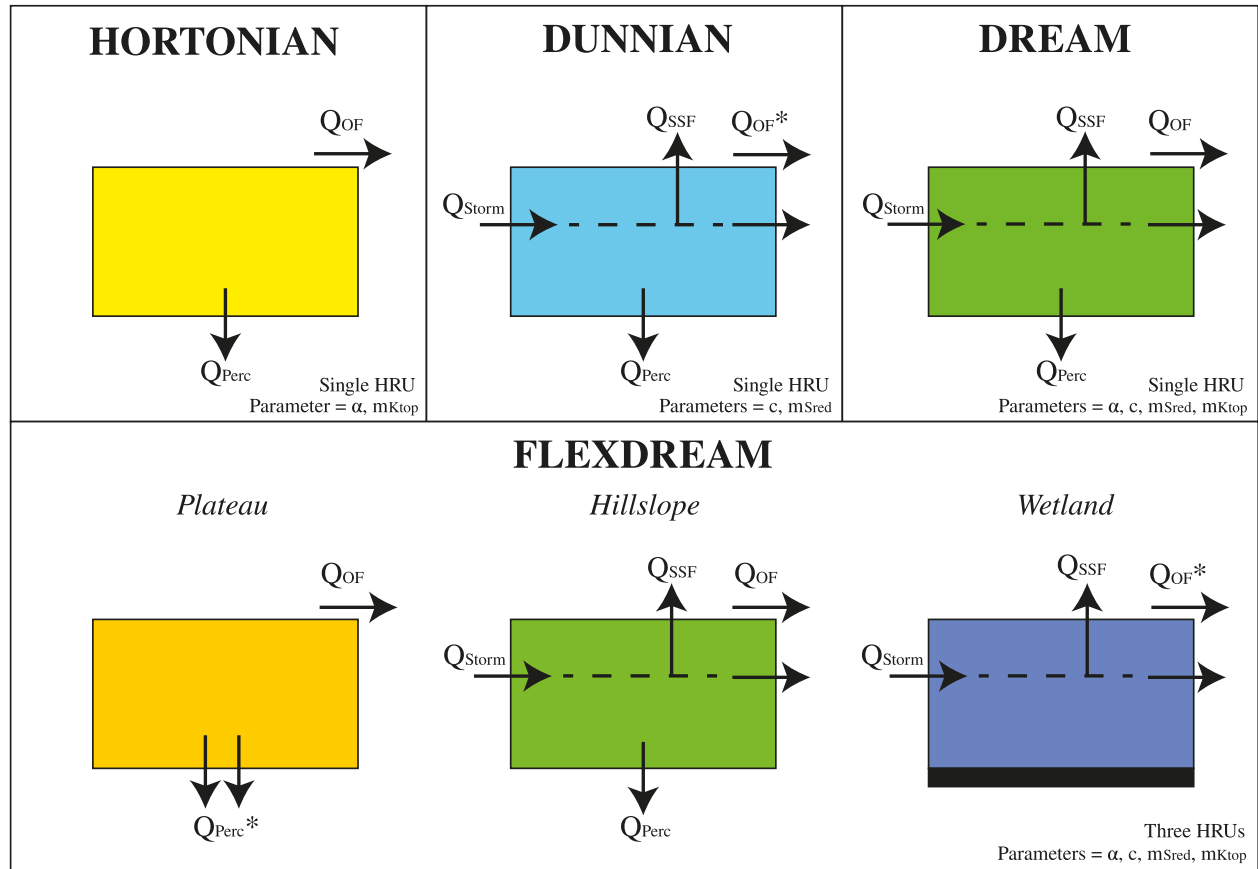


Figure 4. Schematization of the different structures, calibration parameters and processes simulated in each structure. Q_{OF}^* contrary to Q_{OF} does not have Hortonian characteristics, while Q_{Perc}^* contrary to Q_{Perc} can occur if the soil moisture content is higher than the wilting point.

interpretation of the ODEs, while in Figure 4 the main differences of each model configuration are represented more schematically with respect to the sole soil bucket. For a detailed description of all the fluxes, some of which are not covered in this manuscript, readers should refer to Perrini et al. (2024). Of note, the total runoff routed in the hydrodynamic code is intended as the summation of Q_{OF} and Q_{SSF} .

3.2.1. HORTONIAN

The HORTONIAN structure mainly produces HOF across the entire catchment. The infiltration-excess mechanism is triggered when the rainfall rate exceeds a specified threshold (K_{top}). In the absence of this condition, the bucket continues to fill until saturation is reached, resulting eventually in SOF. The mathematical implementation of this process is derived from the modified rational formulation proposed by De Smedt et al. (2000), that can be adapted as follows:

$$P_{Net} = \begin{cases} P_{Eff} \cdot C_p \cdot \left(\frac{S_{Soil}}{S_{Soil,max}} \right)^\alpha, & \text{if } S_{Soil} < S_{Soil,max} \quad \& \quad \text{if } P_{Eff} > K_{top} \\ 0, & \text{if } S_{Soil} < S_{Soil,max} \quad \& \quad \text{if } P_{Eff} \leq K_{top} \\ P_{Eff}, & \text{if } S_{Soil} = S_{Soil,max} \end{cases} \quad (4)$$

$$Q_{Inf} = P_{Eff} - P_{Net} \quad (5)$$

where C_p (–) is the potential runoff coefficient, parametrized as a function of the terrain slope, soil type and macro-classes of land use following Liu and De Smedt (2004), α (–) is the exponent used as calibration parameter. Q_{OF} is later estimated subtracting to the P_{Net} the retention in the surface depressions estimated as in Linsley Jr et al. (1975).

To evaluate the triggering of surface ponding before complete saturation of the soil columns, the saturated hydraulic conductivity of the topsoil (K_s [LT^{-1}]) is generally adopted as a threshold in different hydrological models (see e.g. Francés et al., 2007; Maxwell et al., 2014). To account for the K_s parametrization uncertainties a correction factor (2nd calibration parameter) is considered, namely a simple multiplier ($m_{K_{top}}$) to increase or decrease the K_{top} threshold at the catchment scale starting from the spatial map of K_s .

In wet conditions, also the percolation out of the soil bucket is triggered. This process is assumed to be driven only by gravity and it is computed following Eagleson (1978):

$$Q_{Perc} = \begin{cases} K_{perc} \cdot \left(\frac{S_{Soil}}{S_{Soil,max}} \right)^{\frac{2+3m}{m}}, & \text{if } S_{Soil} > S_{Soil,c} \\ 0, & \text{if } S_{Soil} \leq S_{Soil,c} \end{cases} \quad (6)$$

where a K_{perc} is the exponentially decreased K_s with the depth (in meters) and m the pore-size distribution index estimated as a function of the soil texture and organic content as in Rawls and Brakensiek (1985). Note that the HORTONIAN model computes merely a vertical soil water budget, hence Q_{Storm} and Q_{SSF} in Equations 2 and 3 are excluded from the computation, and the entire S_{Soil} is static storage with respect to the horizontal fluxes.

3.2.2. DUNNIAN

The DUNNIAN model generates SOF and SSF throughout the catchment. It represents a falsification of the HORTONIAN hypothesis and of our prior perceptual understanding of the catchment. Each grid within the domain behaves like a Dunnian bucket, allowing rainfall to infiltrate until soil capacity is exceeded and SOF is triggered. In the DUNNIAN structure the P_{Net} , triggered by the rainfall intensity as consequence of Equation 4, is replaced with:

$$P_{Net} = \begin{cases} 0, & \text{if } S_{Soil} < S_{Soil,max} \\ P_{Eff}, & \text{if } S_{Soil} = S_{Soil,max} \end{cases} \quad (7)$$

and therefore loses the characteristics of a Horton-type overland flow.

Nevertheless, the DUNNIAN structure is governed by a redistribution equation proposed by Manfreda et al. (2005) to simulate SSF. This soil water redistribution process mimics subsurface water dynamics using a threshold mechanism and a reference index of hydrological similarity (ϕ_{red}), namely the Topographic Wetness Index (I_{TWI}) computed as:

$$I_{TWI} = \ln \left(\frac{a}{\tan(\tilde{\alpha})} \right) \quad (8)$$

where a is the drainage area per unit contour length [L], and $\tilde{\alpha}$ the slope.

The grid-based soil water contents, exceeding a certain soil water content threshold (S_{red}), are transferred for each timestep to those cells where the redistribution index is higher, until local exfiltration occurs. This process is modeled within standalone subdomains of the catchment with no exchange of lateral fluxes between each other, namely the subbasins of different Horton-Strahler order (Figure 1b). This means that the subsurface connectivity is guaranteed at the subbasins scale. The redistribution equation can be written as:

$$Q_{\text{Storm},i} = \begin{cases} \left(\frac{\phi_{\text{red},i} \sum_{i=1}^{N(t)} \max[c(S_{\text{Soil},i} - S_{\text{red},i}), 0]}{\sum_{i=1}^{N(t)} \phi_{\text{red},i}} \right) - \max[c(S_{\text{Soil},i} - S_{\text{red},i}), 0], & \text{if } S_{\text{Soil},i} > S_{\text{red},i} \\ \left(\frac{\phi_{\text{red},i} \sum_{i=1}^{N(t)} \max[c(S_{\text{Soil},i} - S_{\text{red},i}), 0]}{\sum_{i=1}^{N(t)} \phi_{\text{red},i}} \right), & \text{if } S_{\text{Soil},i} \leq S_{\text{red},i} \end{cases} \quad (9)$$

where the subscript i indicate any cell of the subbasin and $N(t)$ is the number of cells that at time t exceeds $S_{\text{red},i}$. While c [T^{-1}] is the redistribution coefficient, a calibration parameter that regulates how much water exceeding the S_{red} threshold (i.e., part of the dynamic storage in Figure 3) must be redistributed for each time step within each sub-catchment. For mass conservation c ranges from zero to the unity. Thus, updating the soil water contents, the SSF is described by the constitutive equation:

$$Q_{\text{SSF}} = \begin{cases} Q_{\text{Storm}}, & \text{if } S_{\text{Soil}} = S_{\text{Soil,max}} \\ 0, & \text{if } S_{\text{Soil}} < S_{\text{Soil,max}} \end{cases} \quad (10)$$

Notably, this structure assumes that the active cells for SSF generation encompasses the entire catchment, while the SSF itself is localized in areas with higher I_{TWI} , which directly influences the VCA for SOF.

In the original version of the DUNNIAN model (Manfreda et al., 2005) the S_{red} was a fixed value corresponding to the soil water content at field capacity ($S_{\text{Soil},c}$). Recognizing the uncertainties in soil parametrization but conserving the philosophy of parsimonious parameter selection, we added here a simple multiplier (m_{Sred}) as a 2nd calibration parameter of the DUNNIAN model to regulate the redistribution initialization, as:

$$S_{\text{red}} = \max(S_{\text{Soil,wp}}, \min(m_{\text{Sred}} \cdot S_{\text{Soil},c}, S_{\text{Soil,max}})) \quad (11)$$

where the $S_{\text{Soil,wp}}$ is the lower limits for the activation of the mechanisms (i.e., static storage in Figure 3).

This kind of abstraction tries to conceptually emulate the water dynamic through the soil-matrix using ϕ_{red} as a descriptor of the preferential flow paths. While replacing the $S_{\text{Soil},c}$ with a calibrated value (generally decreased) seeks to emulate various others minor and fuzzy subsurface processes, as for instance, the macro-pore flow, pipe-flow, the fill and spill flow behavior, fairly independent from the continuum soil matrix properties and yet difficult to simulate rationally. For the DUNNIAN structure, deep percolation occurs in the same conditions of Equation 6.

3.2.3. DREAM

The third hypothesis considers the potential concomitance of HOF, SOF and SSF processes in the catchment. This can be done combining in a single model structure the abovementioned computations of the Q_{OF} in Equation 4 and the Q_{SSF} in Equations 9 and 10.

This structure follows the last version of the DREAM model proposed in Perrini et al. (2024), which computes in cascade surface runoff processes first, followed by subsurface runoff, depending on rainfall intensity and soil saturation, but sharing the same soil bucket. The DREAM uses entirely the ODEs reported in Equations 1–3, and its calibration parameters are both the ones of the HORTONIAN and of the DUNNIAN: α , c , m_{Sred} and m_{Ktop} .

3.2.4. FLEXDREAM

This structure combines the DREAM framework with the FLEX-topo model proposed by Savenije (2010), referred to as FLEXDREAM. FLEX-topo is based on the concept that topographical features influence dominant runoff generation processes, differentiating major landscape units with distinct hydrological behaviors. This

approach leads to models that not only use distributed parameters but also have distributed structures, based on the presence, absence, or extent of the identified landscape units.

The original perceptual model, developed for central European landscapes, included three primary landscape classes—wetlands, hillslopes, and plateaus—identified using the Height Above Nearest Drainage (HAND) index (Rennó et al., 2008) and local slope. Given that the HAND index is sensitive to stream network extraction, the reference streamlines were obtained from the Apulian Civil Protection, as they offer the most accurate representation of real streams, effectively bypassing the limitations of automated threshold-based extraction from Digital Terrain Models. FLEX-topo can be implemented using the Hydrological Response Units (HRU) concept (Leavesley, 1984), resulting in semi-distributed models with a limited number of HRUs (three in the original case) and few state variables. This approach is less suitable for the present study, where the fine temporal resolution needed to capture localized HOF processes requires a more detailed spatial discretization.

In this application, the FLEX-topo concept is adapted to a grid-based approach. This allows for the integration of detailed spatial data—such as digital elevation models, land use maps, soil texture and hydraulic properties, and vegetation parameters—into each grid cell, providing a finer level of spatial resolution. The FLEXDREAM uses the HAND-based landscape classification (Gharari et al., 2011; Nobre et al., 2011) to account for dominant runoff generation sources for each macro-landscape according to the FLEX-topo hypothesis (Figure 1c). Based on experience in other catchments (Fenicia & McDonnell, 2022; Gharari et al., 2011), locations with HAND < 10 m were treated as wetlands, locations with HAND > 10 m and slope < 0.1 were classified as plateau and locations with HAND > 10 m and slope > 0.1 were regarded as hillslopes.

Hence, these various landscapes are associated with individual grid-based model architecture (Figure 4). Following Savenije (2010), in the plateau areas the deep percolation is the main hydrological process and the HOF in occurrence of high rainfall intensities is its supporting process. In hillslope the dominant hydrological process is generally the SSF, but according to Savenije (2010) in arid climates with extensive eroded crop surfaces, a potential supporting mechanism on hillslopes is the HOF. In wetland the dominant process is the SOF with supporting mechanisms the groundwater flow, that is, either SSF or baseflow (here neglected). Here, percolation out of the soil substrate is discouraged due to the water table potentially close to the ground surface. Therefore, we choose to implement for the plateau areas the HORTONIAN structure with no restriction for the percolation on soil wet condition (Q_{Perc}^* in Figure 4), for the hillslope areas the full DREAM, while for the wetland the DUNNIAN omitting the percolation process.

As the FLEX-topo implicitly uses the HAND index to describe hydrological similarity, the I_{TWI} (Figure 1e) is replaced with a HAND-based index (i.e., I_{HAND}) for the “constrained” redistribution mechanism that here occurs only on hillslope and wetland. Unlike the DUNNIAN and DREAM hypotheses, where the SSF generation encompassed each grid cell of the catchment, the FLEXDREAM excludes the entire plateau landscape. To make the I_{HAND} (Figure 1f) comparable to the I_{TWI} , namely high values for the expected saturated areas and low values for dryer areas, we rescaled and inverted the values with a natural logarithm as follows:

$$I_{HAND} = \max(\ln(HAND)) - \ln(HAND) \quad (12)$$

I_{HAND} attains high values in areas near the stream, indicating that the VCA of SSF and SOF are primarily located in wetlands rather than on hillslopes, which remain active areas for the generation of both the processes. This is a reasonable assumption, as the similarity index used in the redistribution equation is meant to identify the most likely areas of catchment where soil water exfiltration occurs.

Of note, although the FLEXDREAM complicates and somehow constrains the distributed structure of the catchment behavior, it continues to have the same set of calibration parameters of the DREAM.

4. Model Evaluation

4.1. Calibration and Validation Strategy

Calibrating event-based models using a single peak flow hydrograph is relatively simple, as a few well-chosen parameters can replicate the shape of both rising and falling limbs depending on an accurate rainfall forcing.

For event-based hindcasts, a more robust way to understand runoff generation, is to select consecutive events over a short period. In these cases, as the catchment state evolves during multiple events, different dynamics and processes are activated, providing a better test of the model ability to dynamically simulate streamflow variability.

Event based models can strongly be dependent on initial storage conditions, which need to be carefully determined. In this case, each model structure underwent a warm-up period used to estimate both the Antecedent Soil Moisture ($S_{\text{Soil},t=0}$) as well as the soil water capacity ($S_{\text{Soil},\text{max}}$). Taking as reference the satellite-detected I_{SWI} on the 16th of January 2023, the distributed soil water contents were derived as follows:

$$S_{\text{Soil},\text{SWI}} = S_{\text{Soil},\text{wp}} + I_{\text{SWI}} (S_{\text{Soil},c} - S_{\text{Soil},\text{wp}}) \quad (13)$$

Here using a multiplier on the map derived soil depth, we calibrated the $S_{\text{Soil},\text{max}}$ minimizing the absolute-PBIAS between the $S_{\text{Soil},\text{SWI}}$ for each sub-catchment and the $S_{\text{Soil},t=0}$ produced from the different model structures (Table 2). The warm-up period was of approximately 6 months to ensure that the $S_{\text{Soil},t=0}$ was independent from the initial soil moisture specified at the start of the warm-up (Kim et al., 2018), here the $S_{\text{Soil},\text{wp}}$. From this prior assessment the $S_{\text{Soil},\text{max}}$ over the entire Salsola catchment was found to be averagely of 170 mm, in line with other distributed and semi-distributed models of literature (Bouaziz et al., 2021).

To assess the explanatory power of different hypotheses within the entire nested catchment, the transferability of each structure was tested through a multi-stage operational flowchart that involved calibrating a single headwater catchment (i.e., SS17) and validating spatially on both a catchment of the same order (i.e., SP18) and a higher-order catchment (i.e., SS16). This kind of spatial validation provides a stringent test of model performance, generally considered more challenging than the split-sample temporal validation of models (Andréassian et al., 2009; Fenicia et al., 2016).

Recognizing that the hydrological-hydrodynamic framework operates on different processes and that even GPU-enhanced shallow water models are inadequate for scenarios requiring numerous model runs (Bermúdez et al., 2019), we implemented a three-stage evaluation process:

- The first stage of the evaluation consisted of the optimization of each structure using a simplified routing method, namely the flowtime routing detailed in Appendix A. Calibration was conducted through an automated search using the Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al., 1992) by optimizing a predefined meta-objective function, detailed in Section 4.2.
- In the 2nd stage, we used the calibrated hydrological model and substituted the simplified routing with the hydrodynamic model, which simulates the routing process based on the 2D-SWE. The meta-objective function is recomputed without recalibration, which allows to assess differences in model performance attributable to the choice of routing method, namely the structural error embedded in the simplified flowtime scheme.
- Finally, in the 3rd stage, the parameters of the donor catchment (SS17) were transferred to the two recipient catchments (SS16 and SP18) to spatially validate the model structure along with the hydrodynamic routing. The workflow representation of the model evaluation is provided in Figure 5.

It is important to emphasize that the 2D SWE model is employed to achieve a deterministic evaluation of surface water routing. While roughness coefficients can be adjusted to calibrate catchment-scale hydrodynamic models (Sanz-Ramos et al., 2021), this study utilizes standardized roughness coefficients to bypass equifinality in relation to other runoff-related parameters, which will be discussed later (Section 5.3). This approach also facilitates the identification of structural deficiencies in the simplified routing scheme compared to the more rigorous hydraulic method.

4.2. Meta-Objective Function

Since is widely recognized that a single metric does not adequately summarize model performance (Crochemore et al., 2015; Fenicia et al., 2007; Gauch et al., 2023), four objective functions were selected to capture various aspects of the streamflow time series, including high flow dynamics, low flow dynamics, single-hydrograph dynamics, and non-parametric metrics.

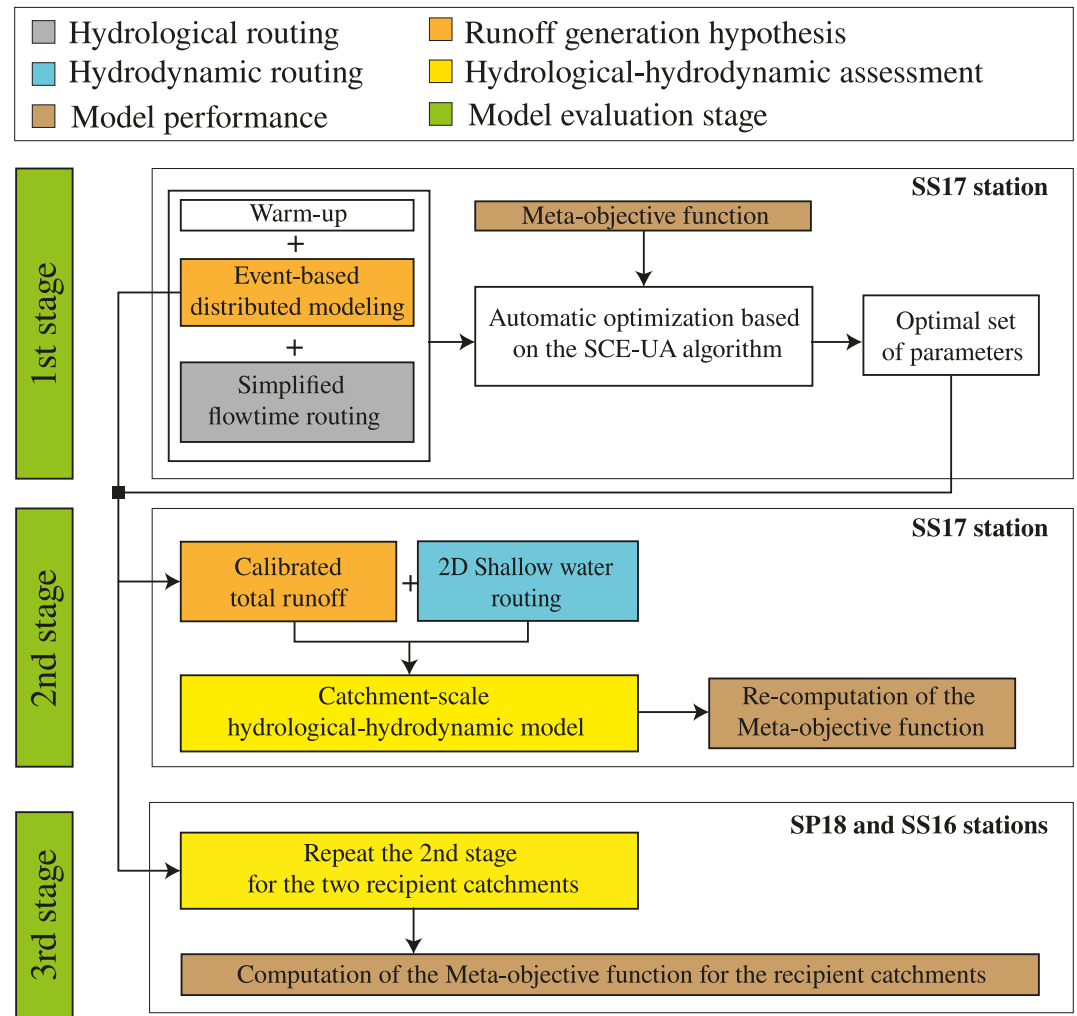


Figure 5. Schematic workflow of the multi-stage operational testing.

These four objective functions were combined into a unique meta-objective function for model optimization, but were assessed individually for model diagnostics. In particular, we used the Kling-Gupta Efficiency (F_{KGE}) and three variants of it, which were averaged with equal weights to compute the meta-objective function ($F_{meta-KGE}$) used for model calibration. Because the time accuracy of the streamflow gauges data was 30 min, the same iteration time step was used for parameter optimization.

4.2.1. KGE

The Kling-Gupta Efficiency (Gupta et al., 2009) is increasingly used in hydrological studies. Its main motivation was to compensate some perceived limitations of the well-known Nash-Sutcliffe Efficiency (F_{NSE}) (Nash & Sutcliffe, 1970), including its tendency to underestimate streamflow variability, its low sensitivity to water balance errors in catchments with highly variable flows, and the inadequacy of using mean flows as a benchmark in such catchments (Schaeffli & Gupta, 2007). The F_{KGE} metric, based on the decomposition of the mean squared error, defines the Euclidean distance computed using the coordinates of bias, standard deviation, and correlation. Its original version is:

$$F_{KGE} = 1 - \sqrt{(1 - \beta)^2 + (1 - \alpha_{KGE})^2 + (1 - r)^2} \quad (14)$$

$$\beta = \mu_{sim}/\mu_{obs} \quad (15)$$

$$\alpha_{KGE} = \sigma_{sim}/\sigma_{obs} \quad (16)$$

$$r = \frac{\sum_{i=1}^n (Q_{obs}(i) - \mu_{obs})(Q_{sim}(i) - \mu_{sim})}{\sqrt{\left(\sum_{i=1}^n (Q_{obs}(i) - \mu_{obs})^2\right)\left(\sum_{i=1}^n (Q_{sim}(i) - \mu_{sim})^2\right)}} \quad (17)$$

where r is the Pearson correlation coefficient, α_{KGE} is the standard deviations (σ) ratio between the simulations (sim) and observations (obs) and β is the ratio of the means (μ).

4.2.2. KGE of Box-Cox Transformed Streamflow

The F_{KGE} , as all the least squares methods, tend to prioritize parameter sets that match hydrological behaviors during high-flow periods. To reduce residual skewness, and focus more on low flows, system-scale metrics include variable transformations such as the log-transform or Box-Cox transform (Pushpalatha et al., 2012). However, the logarithmic transformation can have problems with near-zero flows. Moreover, when used in conjunction with the F_{KGE} , transformations such as logarithmic and Box-Cox generate dependency with respect to the flow unit chosen (e.g., L/s rather than m3/s). To overcome these issues, Santos et al. (2018) developed a peculiar Box-Cox transformation where the streamflow are rescaled adopting the following expression:

$$f_{BC}(Q) = \frac{Q^\lambda - (0.01\mu_{obs})^\lambda}{\lambda} \quad (18)$$

where λ is usually equal to 0.25 for hydrological studies (Vázquez et al., 2008). Therefore the Box-Cox KGE (F_{BCKGE}) is computed starting from the F_{KGE} but replacing the original streamflow timeseries $Q_{obs}(i)$ and $Q_{sim}(i)$ with the transformed variables $f_{BC}(Q_{sim}(i))$ and $f_{BC}(Q_{obs}(i))$.

4.2.3. Non-Parametric KGE

Pool et al. (2018) suggested that calculation of the original F_{KGE} implicitly relies on the assumptions of data linearity, normality, and the absence of outliers. To address these limitations, they reformulated the α_{KGE} and the r terms of the F_{KGE} in a non-parametric form. The Non-parametric KGE (F_{NPKGE}) can be expressed as:

$$F_{NPKGE} = 1 - \sqrt{(1 - \beta)^2 + (1 - \alpha_{NP})^2 + (1 - r_s)^2} \quad (19)$$

The new variability term is built considering the normalized flow duration curve (FDC) to remove the volume information and keep only the distribution signal:

$$\alpha_{NP} = 1 - \frac{1}{2} \sum_{k=1}^n \left| \frac{Q_{sim}(I(k))}{n\mu_{sim}} - \frac{Q_{obs}(J(k))}{n\mu_{obs}} \right| \quad (20)$$

where $I(k)$ and $J(k)$ are the time steps when the k th largest flow occurs within the simulated and observed time series respectively. Introducing the FDC means that the α_{NP} becomes insensitive to errors in the timing and magnitude of individual flood peaks. The Pearson correlation coefficient is replaced with the Spearman rank correlation which generally considers more carefully the mean and low-flow conditions, and can be calculated as follows:

$$r_s = \frac{\sum_{i=1}^n (R_{obs}(i) - \bar{R}_{obs})(R_{sim}(i) - \bar{R}_{sim})}{\sqrt{\left(\sum_{i=1}^n (R_{obs}(i) - \bar{R}_{obs})^2\right)\left(\sum_{i=1}^n (R_{sim}(i) - \bar{R}_{sim})^2\right)}} \quad (21)$$

where R_{obs} and R_{sim} are respectively the rank of the observed and simulated discharge time series and the $\bar{R}$ its means.

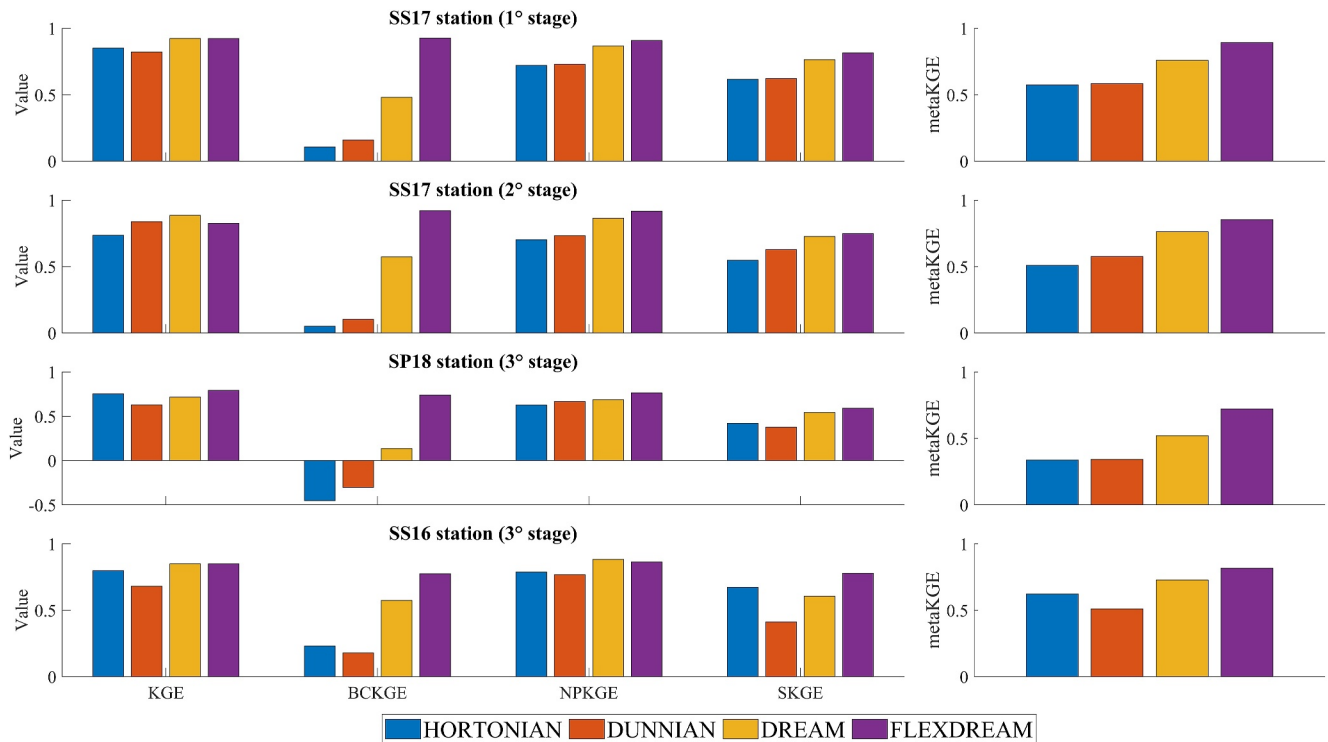


Figure 6. Meta-objective function and individual performance metrics values for each model structure.

4.2.4. Split KGE

The abovementioned metrics rely on full-time span moments statistics, which can obscure especially for event-based analysis the performance of a model over shorter, more variable periods. This can result in an error compensation where discrepancies in different periods might balance each other out masking poor performance in certain stages of the full timespan. This phenomenon is linked to the amalgamation paradox, which describes how statistical relationships can change (increase or decrease) depending on data combinations.

In hydrology, a special case of this statistical paradox has been recently termed “Divide And Measure Non-conformity (DAMN)” since has been proved empirically that F_{NSE} values can be higher than those of individual subsets, and similar effects are also envisaged for the F_{KGE} (Klotz et al., 2024). This aligns also with the need of more nuanced metrics that account for temporal dynamic conditions in hydrological behavior (Gharari et al., 2013). In order to provide a representative assessment of model performance, Fowler et al. (2018) proposed a Split KGE (F_{SKGE}) for long-term hydrological assessments, where the F_{KGE} is computed for each individual year and then averaged.

Here, we follow a similar approach for our event-scale assessments. To better capturing dynamic variations in the multi-event simulation, we chose a fuzzy temporal window that follows the observed flow hydrographs for splitting the F_{KGE} into independent evaluations. Since four clear independent hydrographs were observed at the SS16 outlet, the timespan was divided into four sub-periods (p) to compute first the $F_{KGE}(p)$ for each window and later the averaged F_{SKGE} .

5. Results and Discussions

5.1. Models Performance

Figure 6 illustrates the performance of the four models tested at each stage of the calibration process while the optimal set of parameters obtained with the SCE-UA algorithm is reported in Table 2. In the first stage, the HORTONIAN and DUNNIAN models achieved the lowest and nearly identical $F_{meta-KGE}$ values, whereas FLEXDREAM performed the best. The poorer-performing models exhibited the weakest results in terms of

Table 2

Optimal Parameters, Mean-Absolute-PBIAS, Meta Performance

	α (range 0–3)	m_{Ktop} (range 0–2)	c (range 0–1)	m_{Sred} (range 0.75–1.25)	PBIAS (%)	$F_{\text{meta-KGE}}$ (1st stage) SS17	$F_{\text{meta-KGE}}$ (2nd stage) SS17	$F_{\text{meta-KGE}}$ (3rd stage) SP18	$F_{\text{meta-KGE}}$ (3rd stage) SS16
HORTONIAN	0.35	0.01	–	–	5.53	0.57	0.51	0.34	0.62
DUNNIAN	–	–	0.70	0.75	7.15	0.56	0.57	0.34	0.50
DREAM	0.78	0.34	0.27	0.78	7.83	0.75	0.76	0.52	0.72
FLEXDREAM	0.15	0.33	0.02	0.87	5.25	0.89	0.85	0.72	0.81

F_{BCKGE} , emphasizing deficiencies in low-flow conditions. Although the DREAM and FLEXDREAM models showed similar performance in terms of F_{KGE} , FLEXDREAM outperformed DREAM on other metrics. If evaluated solely based on F_{KGE} , all four model structures would appear to demonstrate comparable skill in fitting the observed hydrographs, with F_{KGE} values always exceeding 0.82. Hence, relying only on F_{KGE} obscures key differences between the models, highlighting the importance of using multiple performance metrics and justifying the selected evaluation criteria.

In the second stage, the calibrated runoff from each model structure in the first stage was transferred to the 2D hydrodynamic model. Here, a slight general decrease in $F_{\text{meta-KGE}}$ was observed for the HORTONIAN and FLEXDREAM models, while the other models showed slight increases. This shift was primarily due to significant changes in F_{KGE} values, caused by overestimation of certain peak flows (see Section 5.2), with the best performances achieved by DREAM and DUNNIAN. Notably, F_{SKGE} , which averages multiple $F_{\text{KGE}}(p)$ values for different events, was highest for FLEXDREAM. In contrast, F_{NPKGE} , which accounts for flow duration curves, remained consistent with first-stage results, indicating that it assigns similar weight to different streamflow magnitudes. Still, FLEXDREAM retained the highest F_{BCKGE} and $F_{\text{meta-KGE}}$.

In the third stage, we validated the models spatially by transferring the calibrated parameters from the SS17 donor catchment to two recipient catchments: SP18 (a similar-sized catchment) and SS16 (a larger catchment) (Figure 8). Unlike traditional regionalization techniques that identify patterns after independently calibrating each catchment, our approach accounts for heterogeneous data layers (e.g., precipitation, land use, soil moisture, vegetation, soil texture) and assumes parameter similarity across the fully nested catchment. This approach allows us to assess the structural validity of each hypothesis by simulating runoff processes calibrated in the first evaluation stage. In the SP18 catchment, the HORTONIAN and DUNNIAN models produced the lowest $F_{\text{meta-KGE}}$ values, this time also with negative F_{BCKGE} scores. Furthermore, despite the F_{KGE} values are above 0.60, all the structures produced much lower F_{SKGE} performance always below 0.60, due to under- and over-estimations of flow peaks across different events. In contrast, at the SS16 catchment, all performance metrics were relatively high, including F_{SKGE} and F_{BCKGE} which had previously been the lowest metrics. This suggests that many least-squares-based metrics remain dependent on flow magnitude, with higher flows generally yielding better performance. Nevertheless, as observed in the second stage, HORTONIAN had a higher F_{SKGE} than DREAM, despite the opposite ranking in F_{KGE} . Once again, FLEXDREAM produced the highest $F_{\text{meta-KGE}}$ scores in both recipient catchments.

Overall, some observations can be drawn from multiple hypothesis testing based on meta-performance evaluation.

- FLEXDREAM consistently outperformed the other models in F_{BCKGE} , indicating that the other three models struggled to simultaneously reproduce both peak and recession portions of the hydrograph, likely due to unresolved structural issues.
- The DUNNIAN model proved inadequate, yielding the lowest average $F_{\text{meta-KGE}}$ in the transferability test (Table 2), reinforcing the need to reject this hypothesis for the nested catchment under investigation.
- Although some $F_{\text{KGE}}(p)$ values spuriously exceeded the aggregated F_{KGE} , F_{SKGE} values were lower than F_{KGE} across all evaluation stages and model structures. This confirms the anticipated DAMN-like behavior of F_{KGE} .

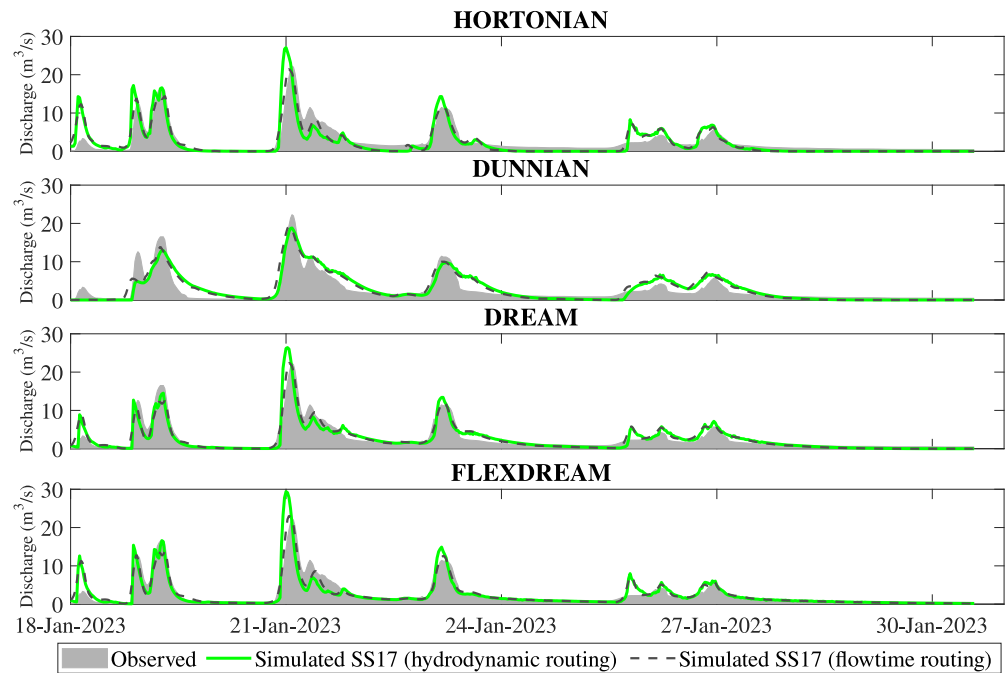


Figure 7. Observed and simulated hydrographs at SS17 station (donor catchment) with both flowtime and hydrodynamic routing.

- Within the context of multiple working hypotheses, DAMN can alter the ranking of the best models (e.g., as seen in the second stage, Figure 6), potentially misrepresenting model fitness. For event-scale modeling, we strongly recommend using F_{SKGE} over the traditional F_{KGE} .

5.2. Consequences of the Routing Methods and the Value of Meta-Metric

The impact of surface routing via strict momentum conservation on simulated streamflow, and consequently on model performance, is not negligible at the SS17 station (Figure 7). Despite identical surface roughness parameterization (i.e., Manning coefficients) in both routing schemes, replacing flowtime routing with the 2D-SWE leads to an overestimation of peak flows and a shorter time-to-peak.

While the flowtime routing scheme assumes time-invariant flux velocities, like most convolutional routing methods, the unsteady nature of surface wave propagation becomes significantly more influential at higher flows when computing depth-dependent flow velocities in 2D-SWE. Furthermore, whereas the Iber+ model employs a refined, non-uniform mesh resolution to simulate the hydrodynamics of main channels accurately, the flowtime routing approach maintains a uniform 50 m spatial discretization. Upscaling the reference topography generally reduces slopes, which, in turn, lowers the static flow velocities calculated by the flowtime scheme.

Such effects have already been recognized in continental-scale studies, where simulated peak discharges and time-to-peaks vary significantly depending on the routing scheme used (Mizukami et al., 2016, 2021; F. Zhao et al., 2017). These findings have critical implications for parameter inference and model optimization. Parameter search processes often compensate for the limitations of simplified routing by adjusting other fluxes and state variables to enhance streamflow-oriented performance metrics (Khatami et al., 2019). Additionally, routing methods can significantly impact performance metrics focused on high flows, such as F_{KGE} , suggesting that parameter recalibration may be necessary depending on the routing scheme employed (Cortés-Salazar et al., 2023).

Automatic optimization of fully hydrological-hydrodynamic models (García-Alén et al., 2024) could potentially address this issue, but such techniques remain computationally prohibitive as they require thousands of simulations. Since these discrepancies are limited to peak magnitudes and timing, a meta-optimization approach—which considers multiple aspects of the streamflow time series—offers a more balanced method for model

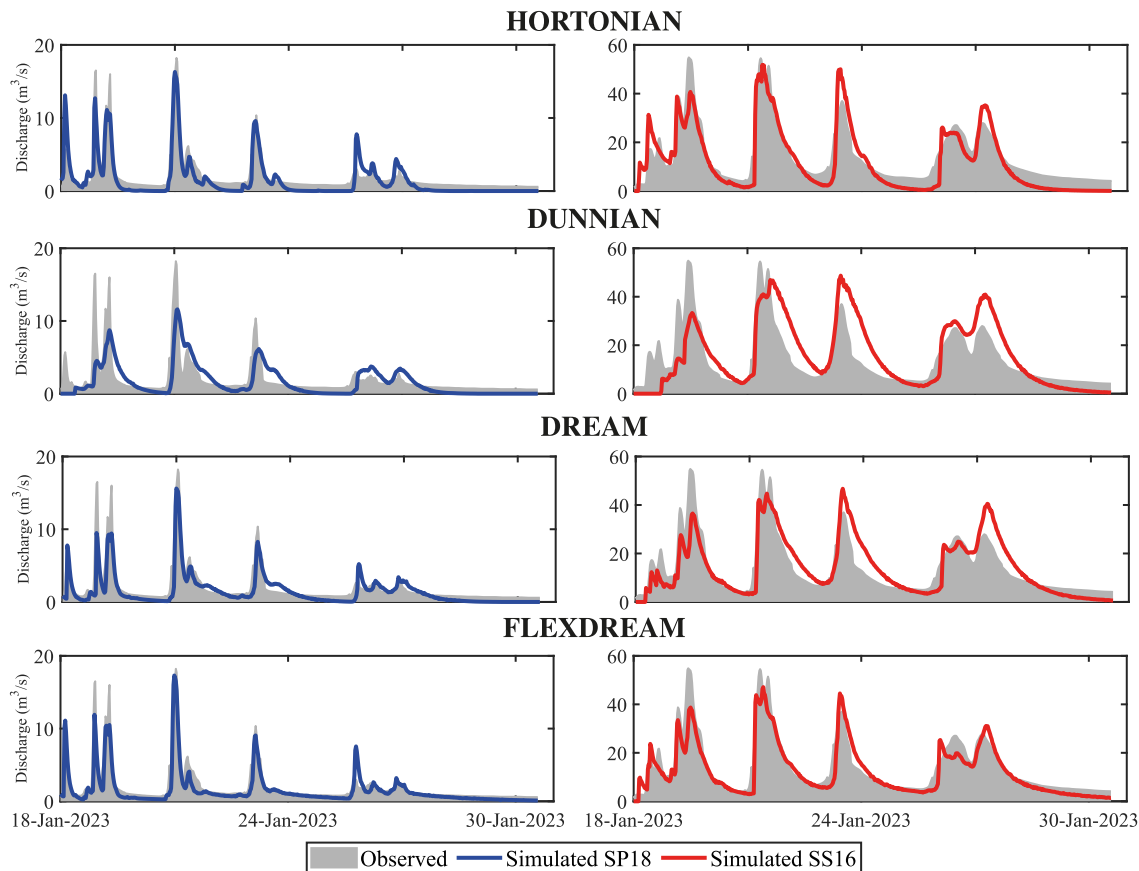


Figure 8. Observed and Simulated hydrographs in the two recipient catchments.

evaluation and comparison, independent of the routing scheme. Supporting this, the FLEXDREAM model consistently achieved the highest $F_{\text{meta-KGE}}$ across all three evaluation stages.

However, automatic calibration of hydrological-hydrodynamic models does not guarantee improved results or success in spatial transferability tests. Unlike at SS17, employing hydrodynamic routing at SS16 and SP18 resulted in an underestimation of peak discharges, particularly for events that were overestimated in the second stage. This suggests unresolved structural issues in the tested hypotheses—not only related to routing processes and runoff generation but also stemming from uncertainties in input data, such as rainfall measurements, spatial rainfall distribution, and rating curves for gauged river stations (McMillan, Krueger, et al., 2012; Renard et al., 2011).

Despite these challenges, spatial transferability performance remained high. Since the study was not intended to develop an operational model for the area, no further parameter optimization was pursued. It is worth noting that a full hydrodynamic simulation of 15 days over the entire Salsola catchment was completed in just 30 min (720 times faster than real time) using an HPC cluster equipped with an NVIDIA A100 GPU (40 GB). This demonstrates the potential of the Iber+ model for real-time flood forecasting and designing storm events with assigned probabilities of occurrence. Although flood inundation modeling was not the primary focus of this study, the presented framework inherently possesses such capabilities.

Interestingly, the DUNNIAN model was the only hypothesis that produced similar peak magnitudes and streamflow shapes regardless of the routing method used. This outcome aligns with theoretical expectations: HOF has short travel times and is more sensitive to routing methods, whereas SSF involves much longer travel times and is less affected by surface routing. An important implication of this analysis is that in catchments lacking HOF, highly specific routing approaches may not be essential for achieving accurate flow simulations. What

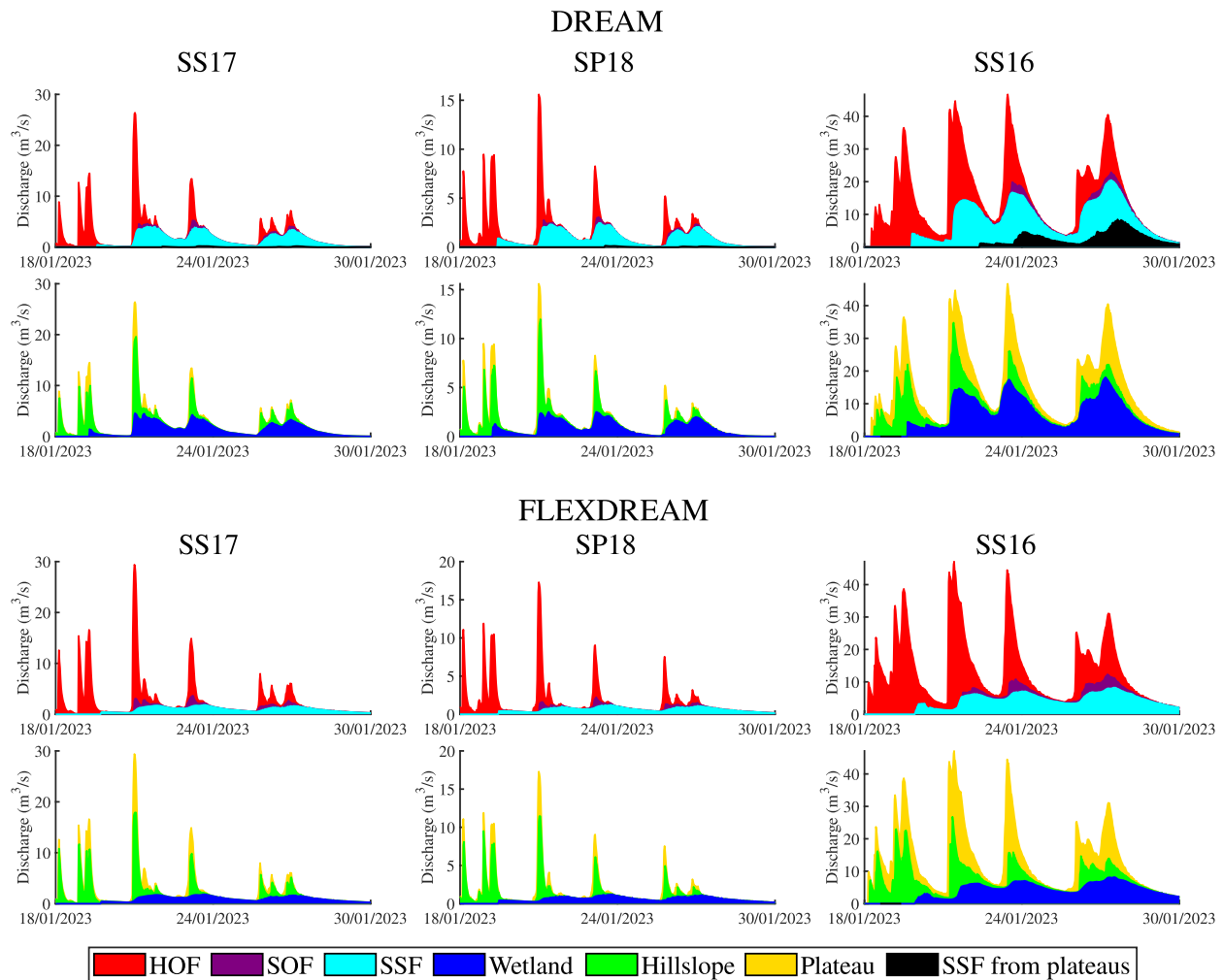


Figure 9. Surface-subsurface and landscape-based streamflow partitioning referred to the DREAM and FLEXDREAM structures. Acronyms in the legend are detailed in Section 1.1.

remains clear is that routing schemes in event-based hydrological applications must be carefully selected, as they represent a significant source of model uncertainty.

5.3. Parameters Interpretability and Processes Realism

As part of the post-model diagnostic, we examined the significance of the optimal parameter sets and the spatial consistency of SSF and SOF source areas. Models incorporating HOF achieved their best performances by significantly reducing the $m_{K_{top}}$ and setting the α to values lower than the unity (Table 2). These findings may be linked to “soil water repellency”, a phenomenon common in arid regions, which can drastically reduce infiltrability and contradict the typical assumption that dry soils have higher infiltration rates than wet soils (Cerdà & Doerr, 2007; Doerr et al., 2003). Indeed, the calibrated values suggest that Horton-type processes activate much earlier than expected based on reference topsoil saturated hydraulic conductivity, and infiltration rates are weakly dependent on the degree of soil saturation. The m_{Sred} SSF-related parameter, indicates that horizontal fluxes initiate before field capacity is reached. Instead, c values are less interpretable, as they depend on the spatial distribution of the similarity index used to emulate subsurface flow dynamics.

Focusing on the two best-performing models, DREAM and FLEXDREAM, we further analyzed the spatial variability of surface and subsurface runoff partitioning alongside the HAND-based landscape classification (Figure 9). In FLEXDREAM, SOF and SSF are predominantly generated in wetland areas rather than on hillslopes, with streamflow contributions from wetlands and SSF closely aligned. In contrast, DREAM generates

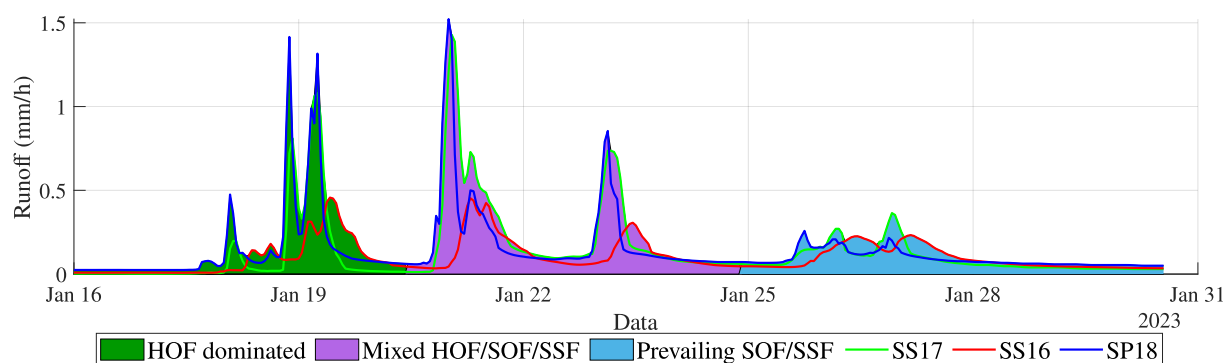


Figure 10. Disentangled dominant runoff generation sources. Acronyms in the legend are detailed in Section 1.1.

SSF across the entire catchment, with a notable portion originating from plateau regions at the SS16 outlet. This process representation in DREAM appears physically inconsistent, as following Savenije (2010), subsurface runoff and saturation-excess processes should be minimal in plateau areas.

Unlike FLEXDREAM, which differentiates macro-landscapes, DREAM employs I_{TWI} to model soil water redistribution at the sub-basin scale. I_{TWI} can yield high values in low-contributing areas with gentle slopes, erroneously identifying upland areas as potential wet regions. This issue is avoided in FLEXDREAM, where subsurface dynamics are restricted to wetlands and hillslopes, and would have been heavily discouraged even without constrained redistribution due to the use of I_{HAND} as the similarity index. Although I_{TWI} is a standardized and widely used hydrological similarity index, its explanatory power in describing saturated areas is generally limited to wet regions (Ali et al., 2014; Güntner et al., 2004). Indeed, previous studies have already shown that HAND-based models provide better representations of saturated areas than TWI-based models in various case studies (e.g., Gao et al., 2019).

The FLEXDREAM structure simulates internal catchment processes in a way that is consistent with the conceptual understanding of landscape behavior, reinforcing the idea that its superior performance is rooted in more accurate modeling assumptions. The DREAM, however, remains a generalization of the FLEXDREAM, meaning that in certain hillslope dominated landscapes the two hypotheses may be very similar. Despite its advantages, the flexible structure still has room for improvement, as outlined in the next section.

5.4. Event-Based Dominant Runoff Generation Mechanisms

By comparing streamflow patterns from different model structures (Figures 7 and 8), we gain key insights into catchment behavior. Simulated hydrographs show that the HORTONIAN hypothesis effectively captures the timing and magnitude of peak discharges but exhibits rapid recessions to zero flow, particularly after the second storm event. This behavior likely stems from the model's inability to account for slower processes such as SSF. Conversely, the DUNNIAN model, which only considers SOF and SSF, overestimates low flows and underestimates high flows, especially during the first three storm events. Additionally, it completely fails to reproduce the initial small hydrograph peak and distorts the double-peak structure of the subsequent hydrograph, producing a single, broader peak instead. Models incorporating HOF tend to overestimate the initial peak, likely due to uncalibrated initial abstractions (e.g., interception and surface depression storage), which could be refined in future analyses. Overall, the FLEXDREAM and DREAM models, which simulate HOF, SOF, and SSF, provide a good match for both peak magnitudes and low flows, suggesting that both surface and subsurface processes contributed to runoff generation during the four storm events.

Identifying runoff (macro)sources typically requires isotope or geochemical tracers from rainfall and streamflow. Although such data were unavailable in this study, we retrospectively inferred the dominant runoff processes for each storm event (Figure 10) as follows:

- At the first event, the poor performance of the DUNNIAN model, compared to the strong performance of the other models, suggests that SOF and SSF were minimal or absent at this stage. Instead, HOF—driven by high rainfall intensities—appears to dominate (Figure 9). This is further supported by headwater catchment time

series, where high runoff magnitudes coincide with minimal recession volumes (Estrany et al., 2010), indicating that even recessions at the outlet were primarily HOF-driven.

- At the second & third events, a combination of surface and subsurface processes is more likely. This is indicated by increasing recession volumes in the headwaters, with hydrograph tails rising above pre-event levels (see also Figure 2), signaling the activation of slower runoff processes as the catchment became wetter. Nevertheless, high flow magnitudes remained strongly associated with intense rainfall.
- At the fourth event, HOF became marginal. The significant drop in runoff coefficients in headwater catchments and the overestimation of gentle peaks by all models incorporating HOF suggest that the rainfall threshold necessary to trigger infiltration-excess was no longer exceeded. This is further supported by the fact that the DUNNIAN model produced a double-peak hydrograph at SS17, closely matching observations, while all HOF-based models simulated a triple-peak hydrograph. Similarly, at SS16, models with HOF produced an irregular hydrograph responding to rainfall intensity (Figure 8), while the DUNNIAN model, relying solely on SOF and SSF, simulated a smoother double-peak hydrograph, which aligned better with observed streamflow.

These findings suggest that after the first three storm events, as the catchment became wetter and rainfall intensities decreased, soil connectivity pathways across the Salsola Catchment were already active, facilitating SOF and SSF rather than HOF. This aligns with prior understanding, where the runoff coefficient for the fourth event at the outlet increased, initially attributed to a combination of surface and subsurface runoff processes.

Despite recent skepticism in the literature, our analysis strongly reinforces the necessity of modeling both surface and subsurface processes in semi-arid regions, even at the event scale. While the alternation and evolution of multiple runoff-generating processes are complex and challenging to capture through modeling alone, considering them is essential to achieving reliable results over consecutive events. Future research should apply this hypothesis-testing approach in nested catchments, where hydrochemical data and saturated area detection are also available to cross-validate the dominant runoff mechanisms. Nevertheless, by combining and falsifying hypotheses, we successfully disentangled key processes through model outcomes and streamflow signatures.

For future modelers, these considerations on how to further develop and improve FLEXDREAM should be kept in mind:

- Refine the K_{top} threshold to reduce HOF occurrences as the saturation ratio increases, aligning with the soil water repellency phenomenon described earlier.
- Enhance FLEXDREAM's representation of saturation areas and subsurface processes by incorporating a more dynamic activation/deactivation of contributing areas.
- Account for plateau connectivity to wetlands, as subsurface connectivity increases under wet conditions, linking larger portions of the catchment. During dry conditions, reduced connectivity limits flow transmission and lowers flow volumes (Fenicia et al., 2014; Jencso et al., 2009).
- Distinguish between uplands and terraces, despite their shared HAND-based landscape classification, as they may exhibit different hydrological behaviors.

6. Conclusion

This study highlights the explanatory power of multiple working hypotheses in characterizing dominant runoff mechanisms at the event scale. To address the structural limitations of lumped catchment-scale models, we integrated bucket-type hydrological models with a 2D hydrodynamic model, aiming to balance model complexity and data availability. By extending the FLEX-topo model to DREAM, we developed a grid-based perceptual structure to test catchment functioning, in combination with the Iber+ 2D-SWE model. Among competing hypotheses, FLEXDREAM produced better results for better reasons, both in terms of performance metrics and internal process consistency.

One of the key findings is that even minor improvements in quantitative metrics can signify substantial progress in embedding more realistic processes into simulated streamflow. This underscores that without optimizing models using meta-performance metrics ($F_{meta} - KGE$), the conclusions could have been markedly different, making it difficult to unequivocally identify the best-performing hypothesis.

For event-based applications, the choice of routing methods proved crucial in influencing peak discharges, emphasizing the need to treat routing as an integral part of model structural uncertainties and evaluation, particularly in Horton-type overland flow-dominated catchments. This highlights the necessity of a balanced

objective function that considers multiple aspects of the streamflow time series, rather than relying solely on least-squares-based performance metrics. Indeed, the well-known F_{KGE} was found to be prone to the DAMN issue, reinforcing the recommendation to use time-consistent metrics such as F_{SKGE} .

Overall, the proposed integrated modeling framework served as a valuable interpretive tool for evaluating surface and subsurface processes that dominate runoff generation under varying conditions. It offers an alternative to existing flexible frameworks, with the capability to test grid-based hydrological hypotheses alongside a 2D hydrodynamic model, making it also well-suited for applications requiring flood inundation estimation. Our findings reinforce the idea that simulating the rainfall-runoff relationship in semi-arid catchments is challenging with a priori assumptions. However, we successfully disentangled a rapid alternation of surface and subsurface processes that drive runoff generation in the nested Salsola catchment. Therefore, event-based hydrological models in semi-arid regions should account for both processes, at least in their prior assumptions.

Ultimately, the research reaffirms two key insights: the fundamental role of landscape in shaping hydrological processes, which should remain central to model development, and the value of event-scale hypothesis testing, which compensates for the lack of additional observations beyond streamflow data.

Appendix A

The surface hydrological routing adopted for a fast calibration uses topographic, topological, and hydrologic information from data layers of the catchment system. It can be categorized as a distributed convolution of the wave propagation. For each grid, a flow path can be defined to the basin outlet, allowing for the calculation of the corresponding flow time under the assumption that the local (cell) velocity is a static characteristic constant over time. Flow directions are assigned using the D8 method based on the steepest downhill slope (O'Callaghan & Mark, 1984). The flow velocity is calculated using Manning's equation:

$$V = n^{-1} R^{-2/3} \tilde{\alpha}^{1/2} \quad (A1)$$

where $\tilde{\alpha}$ is the slope, n is the roughness coefficient assigned in function of the land uses and the R is the hydraulic radius is computed and calibrated as in Manfreda et al. (2005).

The grid-based flow time f_i is determined as the time it takes for water particles to travel along the flow path from the cell to the basin (sub)outlet. The streamflow generated from the total runoff is then calculated as the convolution from all cells, considering their respective f_i , as expressed by the following equation:

$$Q_{\text{Streamflow},t} = \sum_{d=1}^{\tau_{\max}} R_{t-f}(f_i = d) \quad (A2)$$

where $Q_{\text{Streamflow},t}$ represents the streamflow (m^3/s), $\tau_{\max}$ is the maximum flow time in the basin, and d is the index of the summation expressed as a multiple of the time step.

Data Availability Statement

All the data used in the study are mentioned and referenced in the manuscript. The numerical simulations were done with the software Iber v3.2.2, which can be downloaded for free of charge from www.iberaula.com.

References

- Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Tarolli, P. (2014). A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. *Earth Surface Processes and Landforms*, 39(3), 399–413. <https://doi.org/10.1002/esp.3506>
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). FAO Irrigation and drainage paper No. 56. *Food and Agriculture Organization of the United Nations*, 56(97), e156.
- Al-Qurashi, A., McIntyre, N., Wheeler, H., & Unkrich, C. (2008). Application of the Kinos2 rainfall-runoff model to an arid catchment in Oman. *Journal of Hydrology*, 355(1–4), 91–105. <https://doi.org/10.1016/j.jhydrol.2008.03.022>
- Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., et al. (2009). HESS Opinions Crash tests for a standardized evaluation of hydrological models. *Hydrology and Earth System Sciences*, 13(10), 1757–1764. <https://doi.org/10.5194/hess-13-1757-2009>

Acknowledgments

The present study contains part of the results of Ph.D. research by P. Perrini, whose fellowship (CUP H99J21010150001 n DOT20THYKL-4) is supported by PON project: "Programma Operativo Nazionale Ricerca e Innovazione 2014–2020", risorse FSE REACT-EU,—Azione IV.4—Dottorati di ricerca su tematiche dell'innovazione. This research was supported by the financial resources of the "National Centre for HPC, Big Data and Quantum Computing. Area tematica: Simulazioni, calcolo e analisi dei dati ad alte prestazioni PNRR MUR—M4C2—I 1.4"; and of the "Operational Agreement between the Autorità di Bacino Distrettuale dell'Appennino Meridionale (ABDAM) and the Consorzio Interuniversitario per l'Idrologia (CINID)". The first author would like to thank Beatrice Lioi of the Apulian Civil Protection for providing the hydrometric and meteorological data, and Francesco Chiaravallotti of the National Research Council for supplying the rainfall-radar products utilized in this study.

- Antonetti, M., Buss, R., Scherrer, S., Margreth, M., & Zappa, M. (2016). Mapping dominant runoff processes: An evaluation of different approaches using similarity measures and synthetic runoff simulations. *Hydrology and Earth System Sciences*, 20(7), 2929–2945. <https://doi.org/10.5194/hess-20-2929-2016>
- Astagneau, P. C., Bourgin, F., Andréassian, V., & Perrin, C. (2022). Catchment response to intense rainfall: Evaluating modelling hypotheses. *Hydrological Processes*, 36(8), e14676. <https://doi.org/10.1002/hyp.14676>
- Atkinson, S. E., Woods, R. A., & Sivapalan, M. (2002). Climate and landscape controls on water balance model complexity over changing timescales. *Water Resources Research*, 38(12), 50–51. <https://doi.org/10.1029/2002wr001487>
- Barbero, G., Costabile, P., Costanzo, C., Ferraro, D., & Petaccia, G. (2022). 2D hydrodynamic approach supporting evaluations of hydrological response in small watersheds: Implications for lag time estimation. *Journal of Hydrology*, 610(November 2021), 127870. <https://doi.org/10.1016/j.jhydrol.2022.127870>
- Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., et al. (2018). Soil moisture from fusion of scatterometer and SAR: Closing the scale gap with temporal filtering. *Remote Sensing*, 10(7), 1030. <https://doi.org/10.3390/rs10071030>
- Bermúdez, M., Cea, L., & Puertas, J. (2019). A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. *Journal of Flood Risk Management*, 12(S1), e12522. <https://doi.org/10.1111/jfr3.12522>
- Beven, K. J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *Wiley Interdisciplinary Reviews: Water*, 5(3), e1278. <https://doi.org/10.1002/wat2.1278>
- Beven, K. J., Lamb, R., Kirkby, M. J., & Freer, J. E. (2020). A history of TOPMODEL. *Hydrology and Earth System Sciences Discussions*, 1–44.
- Birkel, C., Soulsby, C., & Tetzlaff, D. (2014). Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables. *Water Resources Research*, 50(4), 3481–3501. <https://doi.org/10.1002/2013wr014925>
- Bladé, E., Cea, L., Corestein, G., Escolano, E., Puertas, J., Vázquez-Cendón, E., et al. (2014). Iber: Herramienta de simulación numérica del flujo en ríos. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, 30(1), 1–10.
- Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., et al. (2021). Behind the scenes of Streamflow Model performance. *Hydrology and Earth System Sciences*, 25(2), 1069–1095. <https://doi.org/10.5194/hess-25-1069-2021>
- Calciandro, A., Lamaddalena, N., Stellati, M., & Steduto, P. (2005). Caratterizzazione agroecologica della Regione Puglia. In *Funzione della potenzialità produttiva: Progetto Acla 2* (p. 179). Opuscolo Divulgativo.
- Camporese, M., Paniconi, C., Putti, M., & McDonnell, J. J. (2019). Fill and spill hillslope runoff representation with a Richards Equation-based model. *Water Resources Research*, 55(11), 8445–8462. <https://doi.org/10.1029/2019wr025726>
- Cea, L., Álvarez, M., & Puertas, J. (2024). Using integrated hydrological–hydraulic modelling and global data sources to analyse the February 2023 floods in the Umbeluzi Catchment (Mozambique). *Natural Hazards and Earth System Sciences*, 24(1), 225–243. <https://doi.org/10.5194/nhess-24-225-2024>
- Cea, L., & Bladé, E. (2015). A simple and efficient unstructured finite volume scheme for solving the shallow water equations in overland flow applications. *Water Resources Research*, 51(7), 5464–5486. <https://doi.org/10.1002/2014wr016547>
- Cerdà, A., & Doerr, S. H. (2007). Soil wettability, runoff and Erodibility of major dry-Mediterranean land use types on calcareous soils. *Hydrological Processes: International Journal*, 21(17), 2325–2336. <https://doi.org/10.1002/hyp.6755>
- Chen, X., Yu, Z., Yi, P., Aldahan, A., Hwang, H.-T., & Sudicky, E. A. (2023). Disentangling runoff generation mechanisms: Combining isotope tracing with integrated surface/subsurface simulation. *Journal of Hydrology*, 617, 129149. <https://doi.org/10.1016/j.jhydrol.2023.129149>
- Chow, V. T. (1959). *Open Channel hydraulics*. McGraw-Hill.
- Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47(9). <https://doi.org/10.1029/2010WR009827>
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, 51(4), 2498–2514. <https://doi.org/10.1002/2015wr017198>
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between Hydrological models. *Water Resources Research*, 44(12). <https://doi.org/10.1029/2007wr006735>
- Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., et al. (2021). The abuse of popular performance metrics in hydrologic modeling. *Water Resources Research*, 57(9), e2020WR029001. <https://doi.org/10.1029/2020wr029001>
- Cortés-Salazar, N., Vázquez, N., Mizukami, N., Mendoza, P. A., & Vargas, X. (2023). To what extent does river routing matter in hydrological modeling? *Hydrology and Earth System Sciences*, 27(19), 3505–3524. <https://doi.org/10.5194/hess-27-3505-2023>
- Costabile, P., Costanzo, C., Kalogiros, J., & Bellos, V. (2023). Toward street-level nowcasting of flash floods impacts based on HPC hydrodynamic modeling at the watershed scale and high-resolution weather radar data. *Water Resources Research*, 59(10), e2023WR034599. <https://doi.org/10.1029/2023wr034599>
- Crochemore, L., Perrin, C., Andréassian, V., Ehret, U., Seibert, S. P., Grimaldi, S., et al. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal*, 60(3), 402–423. <https://doi.org/10.1080/02626667.2014.903331>
- Dal Molin, M., Kavetski, D., & Fenicia, F. (2020). SuperflexPy 1.2. 0: An open source Python framework for building, testing and improving conceptual hydrological models. *Geoscientific Model Development Discussions*, 2020, 1–39.
- de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L., & van Wesemael, B. (2015). A map of the topsoil organic carbon content of Europe generated by a Generalized Additive Model. *European Journal of Soil Science*, 66(1), 121–134. <https://doi.org/10.1111/ejss.12193>
- De Girolamo, A. M., Lo Porto, A., Pappagallo, G., Tzoraki, O., & Gallart, F. (2015). The hydrological status concept: Application at a temporary river (Candelaro, Italy). *River Research and Applications*, 31(7), 892–903. <https://doi.org/10.1002/rra.2786>
- De Smedt, F., Yongbo, L., & Gebremeskel, S. (2000). *Hydrologic modelling on a catchment scale using GIS and remote sensed land use information* (Vol. 45). WIT Transactions on Ecology and the Environment.
- Doerr, S. H., Ferreira, A. J. D., Walsh, R. P. D., Shakesby, R. A., Leighton-Boyce, G., & Coelho, C. O. A. (2003). Soil water repellency as a potential parameter in rainfall-runoff modelling: Experimental evidence at point to catchment scales from Portugal. *Hydrological Processes*, 17(2), 363–377. <https://doi.org/10.1002/hyp.1129>
- Doerr, S. H., Shakesby, R. A., & Walsh, R. (2000). Soil water repellency: Its causes, characteristics and hydro-geomorphological significance. *Earth-Science Reviews*, 51(1–4), 33–65. [https://doi.org/10.1016/S0012-8252\(00\)00011-8](https://doi.org/10.1016/S0012-8252(00)00011-8)
- Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4), 1015–1031. <https://doi.org/10.1029/91wr02985>
- Dunne, T., & Black, R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*, 6(5), 1296–1311. <https://doi.org/10.1029/wr006i005p01296>

- Eagleson, P. S. (1978). Climate, soil, and vegetation: 5. A derived distribution of storm surface runoff. *Water Resources Research*, 14(5), 741–748. <https://doi.org/10.1029/wr014i005p00741>
- Estrany, J., Garcia, C., & Batalla, R. J. (2010). Hydrological response of a small mediterranean agricultural catchment. *Journal of Hydrology*, 380(1), 180–190. <https://doi.org/10.1016/j.jhydrol.2009.10.035>
- Fenicia, F., & Kavetski, D. (2021). Behind every robust result is a robust method: Perspectives from a case study and publication process in hydrological modelling. *Hydrological Processes*, 35(8), e14266. <https://doi.org/10.1002/hyp.14266>
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11). <https://doi.org/10.1029/2010wr010174>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., & Freer, J. (2014). Catchment properties, function, and Conceptual Model representation: Is there a correspondence? *Hydrological Processes*, 28(4), 2451–2467. <https://doi.org/10.1002/hyp.9726>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed Hydrological Models: Analysis of key modeling decisions. *Water Resources Research*, 52(2), 954–989. <https://doi.org/10.1002/2015wr017398>
- Fenicia, F., & McDonnell, J. J. (2022). Modeling streamflow variability at the regional scale: (1) Perceptual Model development through signature analysis. *Journal of Hydrology*, 605, 127287. <https://doi.org/10.1016/j.jhydrol.2021.127287>
- Fenicia, F., Meißner, D., & McDonnell, J. J. (2022). Modeling streamflow variability at the regional scale: (2) Development of a bespoke distributed Conceptual Model. *Journal of Hydrology*, 605, 127286. <https://doi.org/10.1016/j.jhydrol.2021.127286>
- Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2007). A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resources Research*, 43(3). <https://doi.org/10.1029/2006wr005098>
- Fowler, K., Peel, M., Western, A., & Zhang, L. (2018). Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resources Research*, 54(5), 3392–3408. <https://doi.org/10.1029/2017wr022466>
- Francés, F., Vélez, J. I., & Vélez, J. J. (2007). Split-parameter structure for the automatic calibration of distributed Hydrological Models. *Journal of Hydrology*, 332(1–2), 226–240. <https://doi.org/10.1016/j.jhydrol.2006.06.032>
- Gao, H., Birkel, C., Hrachowitz, M., Tetzlaff, D., Soulsby, C., & Savenije, H. H. G. (2019). A simple topography-driven and calibration-free runoff generation module. *Hydrology and Earth System Sciences*, 23(2), 787–809. <https://doi.org/10.5194/hess-23-787-2019>
- Gao, H., Hrachowitz, M., Sriwongsitanon, N., Fenicia, F., Gharari, S., & Savenije, H. H. G. (2016). Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region. *Water Resources Research*, 52(10), 7999–8022. <https://doi.org/10.1002/2016wr019574>
- García-Alén, G., Montalvo, C., Cea, L., & Puertas, J. (2024). Iber-PEST: Automatic calibration in fully distributed Hydrological Models based on the 2D shallow water equations. *Environmental Modelling & Software*, 177, 106047. <https://doi.org/10.1016/j.envsoft.2024.106047>
- García-Feal, O., González-Cao, J., Gómez-Gesteira, M., Cea, L., Domínguez, J. M., & Formella, A. (2018). An accelerated tool for flood modelling based on Iber. *Water (Switzerland)*, 10(10), 1–23. <https://doi.org/10.3390/w10101459>
- Garzon, L. F. L., Johnson, M. F., Mount, N., & Gomez, H. (2023). Exploring the effects of catchment morphometry on overland flow response to extreme rainfall using a 2D Hydraulic-Hydrological Model (IBER). *Journal of Hydrology*, 627, 130405.
- Gauch, M., Kratzert, F., Gilon, O., Gupta, H., Mai, J., Nearing, G., et al. (2023). In defense of metrics: Metrics sufficiently encode typical human preferences regarding Hydrological Model performance. *Water Resources Research*, 59(6), e2022WR033918. <https://doi.org/10.1029/2022wr033918>
- Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, 18(12), 4839–4859. <https://doi.org/10.5194/hess-18-4839-2014>
- Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2011). Hydrological landscape classification: Investigating the performance of HAND based landscape classifications in a central European Meso-scale catchment. *Hydrology and Earth System Sciences*, 15(11), 3275–3291. <https://doi.org/10.5194/hess-15-3275-2011>
- Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2013). An approach to identify time consistent model parameters: Sub-period calibration. *Hydrology and Earth System Sciences*, 17(1), 149–161. <https://doi.org/10.5194/hess-17-149-2013>
- Goudarzi, S., Milledge, D., & Holden, J. (2023). A Generalized Multistep Dynamic (GMD) topmodel. *Water Resources Research*, 59(1), e2022WR032198. <https://doi.org/10.1029/2022wr032198>
- Grimaldi, S., Petroselli, A., Tauro, F., & Porfiri, M. (2012). Temps de concentration: Un paradoxe dans l'hydrologie moderne. *Hydrological Sciences Journal*, 57(2), 217–228. <https://doi.org/10.1080/02626667.2011.644244>
- Güntner, A., Seibert, J., & Uhlenbrook, S. (2004). Modeling spatial patterns of saturated areas: An evaluation of different terrain indices. *Water Resources Research*, 40(5). <https://doi.org/10.1029/2003wr002864>
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463–477. <https://doi.org/10.5194/hess-18-463-2014>
- Guse, B., Fatichi, S., Gharari, S., & Melsen, L. A. (2021). Advancing process representation in Hydrological Models: Integrating new concepts, knowledge, and data. *Water Resources Research*, 57(11), e2021WR030661. <https://doi.org/10.1029/2021wr030661>
- Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, 14(1), 446–460.
- Huang, P., Li, Z., Chen, J., Li, Q., & Yao, C. (2016). Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. *Journal of Hydrology*, 542, 292–303. <https://doi.org/10.1016/j.jhydrol.2016.09.001>
- Iacobellis, V., Gioia, A., Manfreda, S., & Fiorentino, M. (2011). Flood quantiles estimation based on theoretically derived distributions: Regional analysis in Southern Italy. *Natural Hazards and Earth System Sciences*, 11(3), 673–695. <https://doi.org/10.5194/nhess-11-673-2011>
- Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., & Marshall, L. A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach-and plot-scale understanding to the catchment scale. *Water Resources Research*, 45(4). <https://doi.org/10.1029/2008wr007225>
- Jewell, S. A., & Gaussiat, N. (2015). An assessment of kriging-based rain-gauge-radar merging techniques. *Quarterly Journal of the Royal Meteorological Society*, 141(691), 2300–2313. <https://doi.org/10.1002/qj.2522>
- Kavetski, D., Fenicia, F., & Clark, M. P. (2011). Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resources Research*, 47(5). <https://doi.org/10.1029/2010wr009525>
- Khatami, S., Peel, M. C., Peterson, T. J., & Western, A. W. (2019). Equifinality and flux mapping: A new approach to model evaluation and process representation under uncertainty. *Water Resources Research*, 55(11), 8922–8941. <https://doi.org/10.1029/2018wr023750>
- Kidron, G. J. (2021). Comparing overland flow processes between semiarid and humid regions: Does saturation overland flow take place in semiarid regions? *Journal of Hydrology*, 593, 125624. <https://doi.org/10.1016/j.jhydrol.2020.125624>

- Kim, K. B., Kwon, H.-H., & Han, D. (2018). Exploration of warm-up period in conceptual hydrological modelling. *Journal of Hydrology*, 556, 194–210. <https://doi.org/10.1016/j.jhydrol.2017.11.015>
- Kirchner, J. W. (2024). Characterizing nonlinear, nonstationary, and heterogeneous hydrologic behavior using Ensemble Rainfall–Runoff Analysis (ERRA): Proof of concept. *Hydrology and Earth System Sciences*, 28(19), 4427–4454. <https://doi.org/10.5194/hess-28-4427-2024>
- Kirchner, J. W., Feng, X., Neal, C., & Robson, A. J. (2004). The fine structure of water-quality dynamics: The (high-frequency) wave of the future. *Hydrological Processes*, 18(7), 1353–1359. <https://doi.org/10.1002/hyp.5537>
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13–24. <https://doi.org/10.1080/02626668609491024>
- Klotz, D., Gauch, M., Kratzert, F., Nearing, G., & Zscheischler, J. (2024). The divide and measure nonconformity—how metrics can mislead when we evaluate on different data partitions. *Hydrology and Earth System Sciences*, 28(15), 3665–3673. <https://doi.org/10.5194/hess-28-3665-2024>
- Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: An open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. *Geoscientific Model Development*, 12(6), 2463–2480. <https://doi.org/10.5194/gmd-12-2463-2019>
- Leavesley, G. H. (1984). Precipitation-runoff modeling system: User's manual. *US Department of the Interior*, 83(4238).
- Linsley, R. K. Jr., Kohler, M. A., & Paulhus, J. L. H. (1975). *Hydrology for engineers*.
- Liu, Y. B., & De Smedt, F. (2004). *WetSpa extension, A GIS-based Hydrological Model for flood prediction and water management, documentation and user manual*. Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel.
- Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H., & Zehe, E. (2019). A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation. *Hydrology and Earth System Sciences*, 23(9), 3807–3821. <https://doi.org/10.5194/hess-23-3807-2019>
- Manfreda, S., Fiorentino, M., & Iacobellis, V. (2005). Dream: A distributed model for runoff, evapotranspiration, and antecedent soil moisture simulation. *Advances in Geosciences*, 2, 31–39. <https://doi.org/10.5194/adgeo-2-31-2005>
- Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J., Ferguson, I. M., Ivanov, V., et al. (2014). Surface-Subsurface Model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, 50(2), 1531–1549. <https://doi.org/10.1002/2013wr013725>
- McDonnell, J. J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., et al. (2007). Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. *Water Resources Research*, 43(7). <https://doi.org/10.1029/2006wr005467>
- McDonnell, J. J., Spence, C., Karran, D. J., Van Meerveld, H. J., & Harman, C. J. (2021). Fill-and-spill: A process description of runoff generation at the scale of the beholder. *Water Resources Research*, 57(5), e2020WR027514. <https://doi.org/10.1029/2020wr027514>
- McMillan, H. (2020). Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, 34(6), 1393–1409. <https://doi.org/10.1002/hyp.13632>
- McMillan, H. (2022). A taxonomy of hydrological processes and watershed function. *Hydrological Processes*, 36(3), 1–7. <https://doi.org/10.1002/hyp.14537>
- McMillan, H., Krueger, T., & Freer, J. (2012a). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, 26(26), 4078–4111. <https://doi.org/10.1002/hyp.9384>
- McMillan, H., Tetzlaff, D., Clark, M., & Soulsby, C. (2012b). Do time-variable tracers aid the evaluation of Hydrological Model structure? A multimodel approach. *Water Resources Research*, 48(5). <https://doi.org/10.1029/2011wr011688>
- Michailidi, E. M., Antoniadis, S., Koukouvinos, A., Bacchi, B., & Efstratiadis, A. (2018). Timing the time of concentration: Shedding light on a paradox. *Hydrological Sciences Journal*, 63(5), 721–740. <https://doi.org/10.1080/02626667.2018.1450985>
- Ming, X., Liang, Q., Xia, X., Li, D., & Fowler, H. J. (2020). Real-time flood forecasting based on a high-performance 2-D Hydrodynamic Model and numerical weather predictions. *Water Resources Research*, 56(7), e2019WR025583. <https://doi.org/10.1029/2019wr025583>
- Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., et al. (2021). A vector-based river routing model for earth system models: Parallelization and global applications. *Journal of Advances in Modeling Earth Systems*, 13(6), e2020MS002434. <https://doi.org/10.1029/2020ms002434>
- Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., et al. (2016). MizuRoute version 1: A river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, 9(6), 2223–2238. <https://doi.org/10.5194/gmd-9-2223-2016>
- Mul, M. L., Mutibwa, R. K., Uhlenbrook, S., & Savenije, H. H. G. (2008). Hydrograph separation using hydrochemical tracers in the Makanya catchment, Tanzania. *Physics and Chemistry of the Earth, Parts A/B/C*, 33(1), 151–156. <https://doi.org/10.1016/j.pce.2007.04.015>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., & Saleska, S. (2011). Height above the Nearest Drainage—a hydrologically relevant New Terrain Model. *Journal of Hydrology*, 404(1–2), 13–29. <https://doi.org/10.1016/j.jhydrol.2011.03.051>
- O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 28(3), 323–344. [https://doi.org/10.1016/s0734-189x\(84\)80011-0](https://doi.org/10.1016/s0734-189x(84)80011-0)
- Ochoa-Rodriguez, S., Wang, L. P., Willems, P., & Onof, C. (2019). A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research*, 55(8), 6356–6391. <https://doi.org/10.1029/2018WR023332>
- Panicconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, 51(9), 7090–7129. <https://doi.org/10.1002/2015wr017780>
- Perrini, P., Cea, L., Chiaravalloti, F., Gabriele, S., Manfreda, S., Fiorentino, M., et al. (2024). A runoff-on-grid approach to embed hydrological processes in shallow water models. *Water Resources Research*, 60(7), e2023WR036421. <https://doi.org/10.1029/2023wr036421>
- Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, 46(1). <https://doi.org/10.1029/2009wr008066>
- Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*, 44(12). <https://doi.org/10.1029/2007wr006615>
- Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, 63(13–14), 1941–1953. <https://doi.org/10.1080/02626667.2018.1552002>
- Pushpalatha, R., Perrin, C., Le Moine, N., & Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, 420, 171–182. <https://doi.org/10.1016/j.jhydrol.2011.11.055>
- Rawls, W. J., & Brakensiek, D. L. (1985). Prediction of soil water properties for hydrologic modeling. *Watershed Management in the Eighties*, 293–299.

- Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., & Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47(11). <https://doi.org/10.1029/2011wr010643>
- Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping Terra-Firme rainforest environments in Amazonia. *Remote Sensing of Environment*, 112(9), 3469–3481. <https://doi.org/10.1016/j.rse.2008.03.018>
- Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, 22(8), 4583–4591. <https://doi.org/10.5194/hess-22-4583-2018>
- Sanz-Ramos, M., Bladé, E., González-Escalona, F., Olivares, G., & Aragón-Hernández, J. L. (2021). Interpreting the Manning roughness coefficient in overland flow simulations with coupled hydrological-hydraulic distributed models. *Water*, 13(23), 3433. <https://doi.org/10.3390/w13233433>
- Savage, J. T. S., Bates, P., Freer, J., Neal, J., & Aronica, G. (2016). When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrological Processes*, 30(13), 2014–2032. <https://doi.org/10.1002/hyp.10749>
- Savenije, H. H. G. (2010). HESS Opinions" Topography driven conceptual modelling (FLEX-Topo). *Hydrology and Earth System Sciences*, 14(12), 2681–2692. <https://doi.org/10.5194/hess-14-2681-2010>
- Sayama, T., McDonnell, J. J., Dhakal, A., & Sullivan, K. (2011). How much water can a watershed store? *Hydrological Processes*, 25(25), 3899–3908. <https://doi.org/10.1002/hyp.8288>
- Schaeffli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, 21(15), 2075–2080. <https://doi.org/10.1002/hyp.6825>
- Shen, H., Tolson, B. A., & Mai, J. (2022). Time to update the split-sample approach in hydrological model calibration. *Water Resources Research*, 58(3), e2021WR031523. <https://doi.org/10.1029/2021WR031523>
- Sittner, W. T. (1976). WMO project on intercomparison of conceptual models used in hydrological forecasting. *Hydrological Sciences Journal*, 21(1), 203–213. <https://doi.org/10.1080/02626667609491617>
- Smakhtin, V. U. (2001). Low flow hydrology: A review. *Journal of Hydrology*, 240(3), 147–186. [https://doi.org/10.1016/S0022-1694\(00\)00340-1](https://doi.org/10.1016/S0022-1694(00)00340-1)
- Tao, Z., Li, M., Si, B., & Pratt, D. (2021). Rainfall intensity affects runoff responses in a semi-arid catchment. *Hydrological Processes*, 35(4), e14100. <https://doi.org/10.1002/hyp.14100>
- Vázquez, R. F., Willems, P., & Feyen, J. (2008). Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteria approach. *Hydrological Processes: International Journal*, 22(13), 2159–2179. <https://doi.org/10.1002/hyp.6815>
- Weiler, M., McDonnell, J. J., Tromp-van Meerveld, I., & Uchida, T. (2006). *Subsurface stormflow*. Encyclopedia of Hydrological Sciences.
- Wheeler, H., Sorooshian, S., & Sharma, K. D. (2007). *Hydrological modelling in arid and semi-arid areas*. Cambridge University Press.
- Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., et al. (2015). Towards more systematic perceptual model development: A case study using 3 Luxembourgish catchments. *Hydrological Processes*, 29(12), 2731–2750. <https://doi.org/10.1002/hyp.10393>
- Zanetti, F., Camporese, M., & Botter, G. (2024). How do different runoff generation mechanisms drive stream network dynamics? Insights from physics-based modelling. *Hydrological Processes*, 38(7), e15234. <https://doi.org/10.1002/hyp.15234>
- Zhang, L., Potter, N., Hickel, K., Zhang, Y., & Shao, Q. (2008). Water balance modeling over variable time scales based on the Budyko framework—Model development and testing. *Journal of Hydrology*, 360(1–4), 117–131. <https://doi.org/10.1016/j.jhydrol.2008.07.021>
- Zhao, F., Veldkamp, T. I. E., Frieler, K., Schewe, J., Ostberg, S., Willner, S., et al. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, 12(7), 75003. <https://doi.org/10.1088/1748-9326/aa7250>
- Zhao, R. J. (1977). *Flood forecasting method for humid regions of China*. East China College of Hydraulic Engineering.