

## COMMENTARY

WILEY

# When good signatures go bad: Applying hydrologic signatures in large sample studies

Hilary McMillan<sup>1</sup>  | Gemma Coxon<sup>2</sup>  | Ryoko Araki<sup>1</sup>  | Saskia Salwey<sup>2</sup>  |  
 Christa Kelleher<sup>3</sup>  | Yanchen Zheng<sup>2</sup>  | Wouter Knoben<sup>4</sup>  |  
 Sebastian Gnnann<sup>5</sup>  | Jan Seibert<sup>6</sup>  | Lauren Bolotin<sup>1</sup> 

<sup>1</sup>Department of Geography, San Diego State University, San Diego, California, USA

<sup>2</sup>School of Geographical Sciences, University of Bristol, Bristol, UK

<sup>3</sup>Civil and Environmental Engineering, Lafayette College, Easton, Pennsylvania, USA

<sup>4</sup>Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada

<sup>5</sup>Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

<sup>6</sup>Department of Geography, University of Zurich, Zurich, Switzerland

## Correspondence

Hilary McMillan, Department of Geography, San Diego State University, San Diego, California, USA.  
 Email: [hmcmillan@sdsu.edu](mailto:hmcmillan@sdsu.edu)

## Funding information

NSF Hydrologic Sciences Program; Division of Earth Sciences, Grant/Award Number: 2124923

## Abstract

Hydrologic signatures are quantitative metrics that describe streamflow statistics and dynamics. Signatures have many applications, including assessing habitat suitability and hydrologic alteration, calibrating and evaluating hydrologic models, defining similarity between watersheds and investigating watershed processes. Increasingly, signatures are being used in large sample studies to guide flow management and modelling at continental scales. Using signatures in studies involving 1000s of watersheds brings new challenges as it becomes impractical to examine signature parameters and behaviour in each watershed. For example, we might wish to check that signatures describing flood event characteristics have correctly identified event periods, that signature values have not been biased by data errors, or that human and natural influences on signature values have been correctly interpreted. In this commentary, we draw from our collective experience to present case studies where naïve application of signatures fails to correctly identify streamflow dynamics. These include unusual precipitation or flow regimes, data quality issues, and signature use in human-influenced watersheds. We conclude by providing guidance and recommendations on applying signatures in large sample studies.

## KEYWORDS

Signatures, Metrics, Indices, Large sample hydrology, Errors, Uncertainty

## 1 | INTRODUCTION

Hydrologic signatures are quantitative metrics that describe streamflow statistics and dynamics. Examples include runoff ratio, baseflow index or slope of the flow duration curve. Signatures have many applications such as assessing habitat suitability and hydrologic alteration, evaluating hydrologic models, defining similarity between watersheds and investigating watershed processes (McMillan, 2021). Signature concepts have been extended to characterize soil moisture (Araki et al., 2022), water quality (Ebeling et al., 2021) and other hydrologic quantities.

From earlier uses in small numbers of watersheds, signatures are now frequently calculated in large samples of watersheds spanning national to global scales. This expansion is part of a wider move towards large sample studies that develop datasets and draw conclusions across scales, hydro climates and ecosystems (Addor et al., 2020; Kratzert et al., 2023). Studies that calculate signature values over 100s of gauged watersheds include evaluating national models (Almagro et al., 2021; Coxon et al., 2019; Donnelly et al., 2016; Massmann, 2020; McMillan et al., 2016), selecting model structures (David et al., 2022), interpreting machine learning models (Botterill & McMillan, 2022;

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

© 2023 The Authors. *Hydrological Processes* published by John Wiley & Sons Ltd.

Kratzert et al., 2019), predicting signatures from watershed attributes (Addor et al., 2018; Beck et al., 2015; Grantham et al., 2022; Janssen & Ameli, 2021) and classifying watersheds (Kuentz et al., 2017). However, signature use is challenged by large datasets, as it becomes impractical to check whether the signature quantifies the intended hydrograph property in each watershed. Although difficulties in transferring knowledge between individual samples and large populations are well established (Jelinski & Wu, 1996), these issues have accelerated in hydrologic analysis with more widespread use of large sample datasets. As authors we have encountered these challenges, prompting us to write this commentary to share our experiences and provide guidance on applying hydrologic signatures in large sample studies.

## 1.1 | Challenges in using signatures over large samples

Some signatures such as statistics of flow magnitude or timing are simple to calculate and robust to many errors, uncertainties or unusual flow patterns. Other signatures, such as those requiring identification of storm events or recession periods require more time-series processing and may be less robust. All signatures are impacted by missing data in the time-series, and large sample studies often tolerate some missing data to maximize the number of sites. Signature toolboxes such as TOSSH (Gnann, Coxon, et al., 2021), eflows (Patterson et al., 2020) or Pastas (Collenteur et al., 2019) may reject the time-series if missing data exceed a certain percentage of the record, or interpolate the data for smaller missing portions. Signature values can be biased where missing data are systematic, such as missing data under freezing conditions or under extreme high or low flows.

Large sample studies require users to specify signature parameter values. Parameters change how signatures retrieve information from flow series and affect the resulting signature value. For example, recession analysis signatures are influenced by parameters for recession extraction and fitting (Dralle et al., 2017). Large sample studies often use constant parameter values (Addor et al., 2018; McMillan et al., 2022), but these may be unsuitable for some watersheds. Strategies to select parameter values include using sensitivity analysis to explore how parameter choices affect study conclusions (Tashie et al., 2020), conducting in-depth checks for representative watersheds such as one watershed per climate region where signature parameters are sensitive to climate, or deriving parameters from the flow regime (Stoelzle et al., 2020).

Large sample studies often include watersheds with reservoir storage and releases, abstractions or agricultural water use. In human-impacted watersheds in California, signatures could not be calculated because “the annual hydrograph was extremely different compared with the predicted reference condition. These instances would often lack a seasonal flow pattern that the flow calculator relies on to derive subsequent metrics” (Peek et al., 2022). Hydropower operations cause sudden flow changes that make computations of change rates problematic, invalidating recession signatures (Zmijewski & Wörman, 2016). Solutions include designing signatures to measure human impacts (e.g., Indicators of

Hydrologic Alteration, IHA; Magilligan & Nislow, 2005), or using existing signatures such as high flashiness signifying urban effects (Smith & Smith, 2015).

A related challenge is signature interpretability. In small sample studies, researchers often have prior expertise (or a perceptual model) against which they interpret the signature values. In large sample studies, it is more difficult to determine whether the same signature values correspond with the same hydrologic characteristics. Our case studies include human-influenced watersheds, where it is difficult to distinguish the effects of reservoir operations from differences in natural flow regimes, and soil moisture seasonality, where bimodal distributions have multiple possible causes. In small-sample studies, signatures can be checked for correct interpretation, but in-depth checks are usually not feasible for 1000s of watersheds.

## 1.2 | Aims of the paper

In this paper, we present eight case studies illustrating challenges that arose in our work when calculating hydrologic signatures over large samples of watersheds. These include unusual precipitation regimes, data quality issues and human-influenced watersheds. For each study we discuss the aim, which signatures were calculated, the issues that occurred and the solution, if any. We use the lessons learnt to provide guidance on applying signatures in large sample studies.

## 2 | CASE STUDIES: CHALLENGES APPLYING SIGNATURES IN LARGE SAMPLES OF WATERSHEDS

### 2.1 | Signature issues caused by precipitation and flow regimes

Many signatures make assumptions about watershed precipitation and flow regimes. If the assumptions are invalidated, as shown in these examples, signature values can lead to incorrect interpretations.

#### 2.1.1 | Event identification failed in Arizona watersheds with convective monsoon rainfall

We used signatures to investigate overland flow generated by convective monsoonal rainfall in 21 subcatchments of the San Pedro River in Southern Arizona. We calculated infiltration- and saturation-excess signatures from the TOSSH toolbox (Gnann, Coxon, et al., 2021), including the significance of thresholds in plots of event runoff versus rainfall depth or intensity, and coefficients in regression equations to predict event runoff. These signatures rely on event identification algorithms to calculate event rainfall and flow totals. The algorithms identify rainfall peaks and then find the corresponding runoff, or vice versa (McMillan et al., 2011; Tarasova et al., 2018). The algorithms use parameters such as thresholds for magnitude or separation time between events. Checks

on our rainfall-based algorithm showed that afternoon rainfall on consecutive days was often merged into a single event because the algorithm specified a dry period of 12 h to separate events (Figure 1a). Afternoon rainfall could cause a flow peak overlapping with the following day's rainfall, which would be mistakenly associated. In such watersheds, hourly data and careful parameter choice are needed to avoid conflating separate events. We used hourly data, but large sample studies may use daily values (e.g., McMillan et al., 2022), which would be insufficient to identify convective rainstorms.

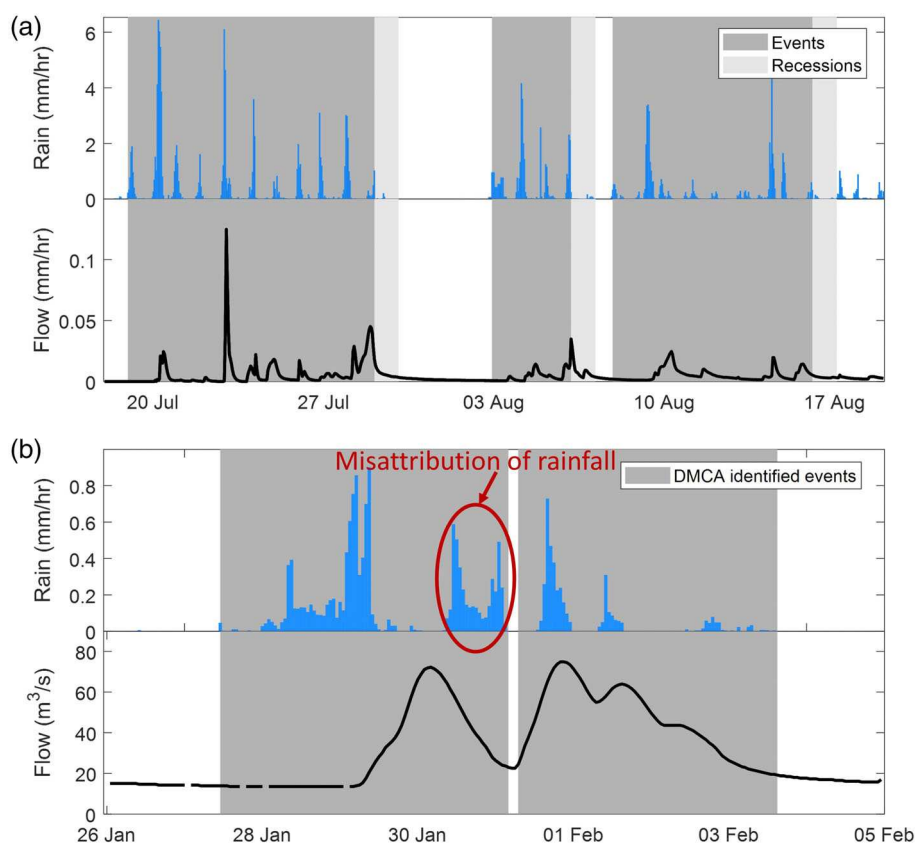
### 2.1.2 | Event identification required manual checks in drizzly United Kingdom

In the United Kingdom (U.K.), most watersheds experience at least 0.2 mm of rainfall every other day, rising to 78% of days for the wettest watersheds. In a large sample U.K. study, we calculated the event runoff coefficient signature (the ratio between the direct runoff volume and event rainfall) to reveal the space–time controls on event characteristics (Zheng et al., 2023a, 2023b). We used flow-based event identification with hourly data (Giani et al., 2022), but short intervals between events sometimes led to the misattribution of rainfall to runoff events (Figure 1b). Given these issues, we used manual checks on 1144 events from the Swale River at Crakehill (NRFA ID 27071) to create event selection criteria. For events with rainfall after the runoff peak, the time interval between events was calculated. If this interval was less than

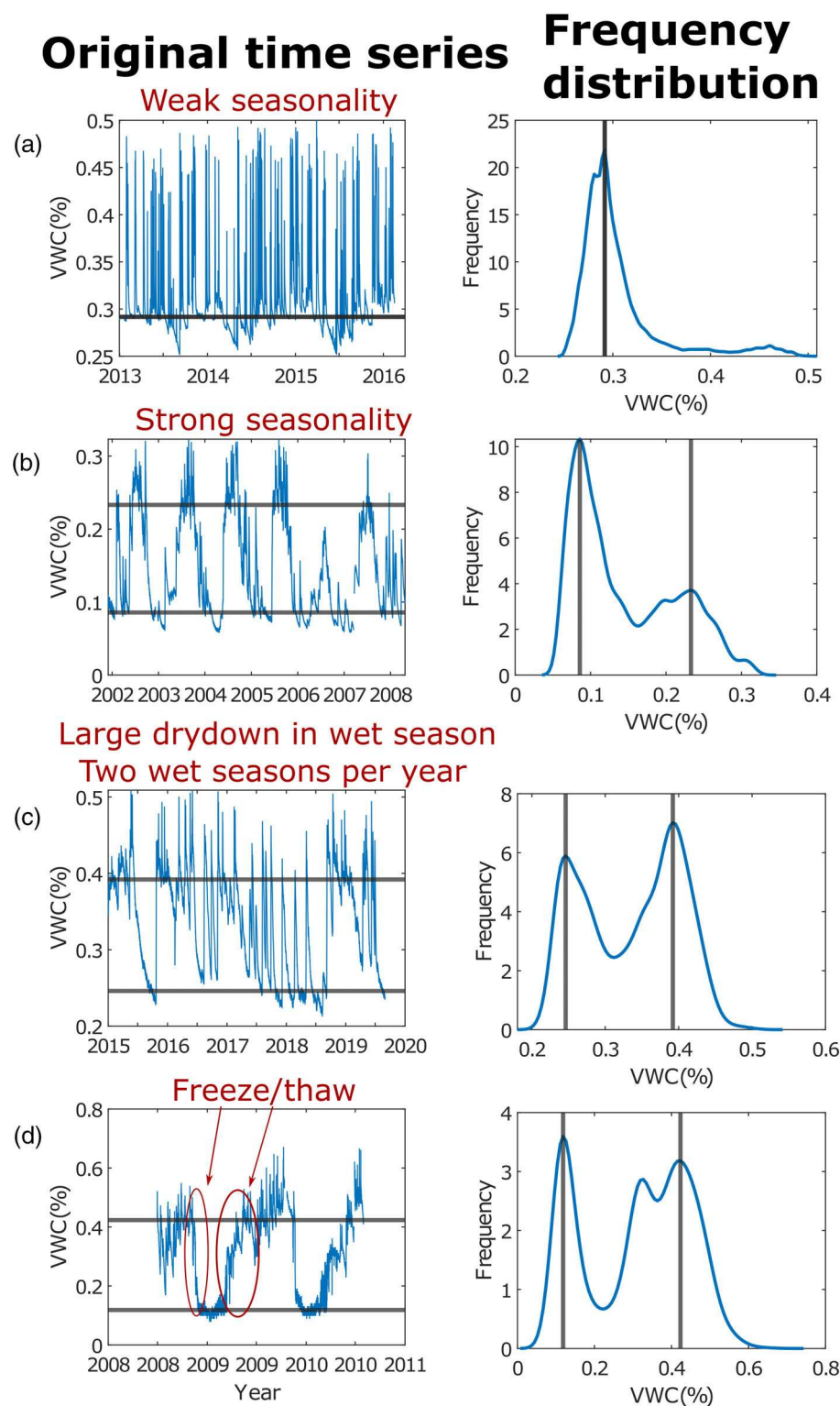
the watershed response time, both events were excluded. These criteria led us to exclude 23 393 out of 903 745 events from 431 watersheds. High exclusion rates occur in large watersheds ( $> 1000 \text{ km}^2$ ) and those with long response times ( $> 1 \text{ d}$ ). In this and the previous example, a more complex algorithm is required to associate rainfall with runoff that allows for the time of concentration.

### 2.1.3 | Soil moisture seasonality was misidentified in sites with multiple wet seasons in a year

We used large sample studies to investigate controls on soil moisture seasonality, particularly the persistent wet and dry seasons that often appear in soil moisture data (Araki et al., 2022, 2023). We used two signature types: the number of peaks of the soil moisture frequency distribution, and (for bimodal distributions only) dates and duration of transitions between wet and dry states. These signatures were developed for a study of 82 New Zealand sites (Branger & McMillan, 2020). The number of frequency peaks was designed to represent the strength of soil moisture seasonality. A unimodal distribution represents weak seasonality under persistently wet or dry climates, whereas a bimodal distribution represents strong seasonality (Sehgal & Mohanty, 2023; Figure 2a,b). However, applications to a wider range of climates showed that bimodal distributions had multiple causes. Sites in Texas showed multiple wet seasons (winter, summer monsoon) and large dry downs during wet seasons due to well-drained soil. This resulted in



**FIGURE 1** Incorrect event identification examples (a) in Arizona, multiple convective rainfall events were incorrectly merged into single storm periods (USGS streamgauge 09471550, San Pedro River Near Tombstone) (b) in U.K., rainfall occurring after the runoff peak was misattributed to the wrong runoff events (NRFA catchment ID 27071, Swale River at Crakehill).



**FIGURE 2** Peaks in the distribution of soil moisture identified under (a) a weakly seasonal climate in Germany, (b) a strongly seasonal climate in Australia, (c) a climate with two wet seasons per year in Texas, (d) a site with soil freeze–thaw in China.

strongly bimodal signatures despite weak seasonality (Figure 2c), and invalidated the algorithm for transition signatures that assumed only one wet season per year. Pseudo-seasonality also occurred in Maqu watershed, China, where soil freeze–thaw created a bimodal distribution unrelated to wet and dry seasons (Figure 2d). Manual checks of rainfall and soil temperature seasonality were needed when interpreting signature values in these different regimes.

## 2.2 | Signature issues caused by data quality

Signatures vary in their robustness to error types (Westerberg & McMillan, 2015). In large sample studies, data errors will vary site-to-site. In our examples, data quality motivated exclusion of data from specific sites or regions, or development of new processing steps to mitigate errors.

### 2.2.1 | Issues with catchment area affects flow magnitude signatures

Catchment polygons describe the area upstream of a flow gauge, and may contain errors due to incorrect outlet coordinates (e.g., Arheimer et al., 2020), errors in catchment delineation method or elevation data errors. Inaccurate basin areas can have large impacts. Streamflow is often transformed from raw units [ $M^3T^{-1}$ ] to specific discharge [ $MT^{-1}$ ], by dividing by basin area. Area errors will lead to incorrect specific discharge, and propagate into signatures. Further errors will occur if meteorological data are snipped to incorrect basin polygons, and used in signatures such as runoff ratio ( $Q/P$ ) and streamflow elasticity ( $\Delta Q/\Delta P$ ). We investigated the impacts of basin polygon errors in the CAMELS dataset of 671 U.S. basins (Addor et al., 2017; Newman et al., 2015). CAMELS includes streamflow observations in [ $feet^3 s^{-1}$ ] and two different area estimates for each basin from GAGES II and Geospatial fabric datasets. We used the two area values to calculate three magnitude-based signatures (5th/95th percentile flows, daily mean flow) in [ $mm d^{-1}$ ]. The differences are typically within  $\pm 5\%$ , but sometimes substantial (Figure 3a–c). The resulting differences in annual streamflow exceed 100 mm for many basins (Figure 3d). Some CAMELS studies exclude watersheds with uncertain area (Knoben et al., 2020; Kratzert et al., 2019) but others use all or most watersheds, implicitly accepting this uncertainty (Addor et al., 2018; Tyralis et al., 2021). Any large sample study using magnitude-based signatures is susceptible to catchment area uncertainty, such as the event runoff coefficient study described previously (Zheng et al., 2023a).

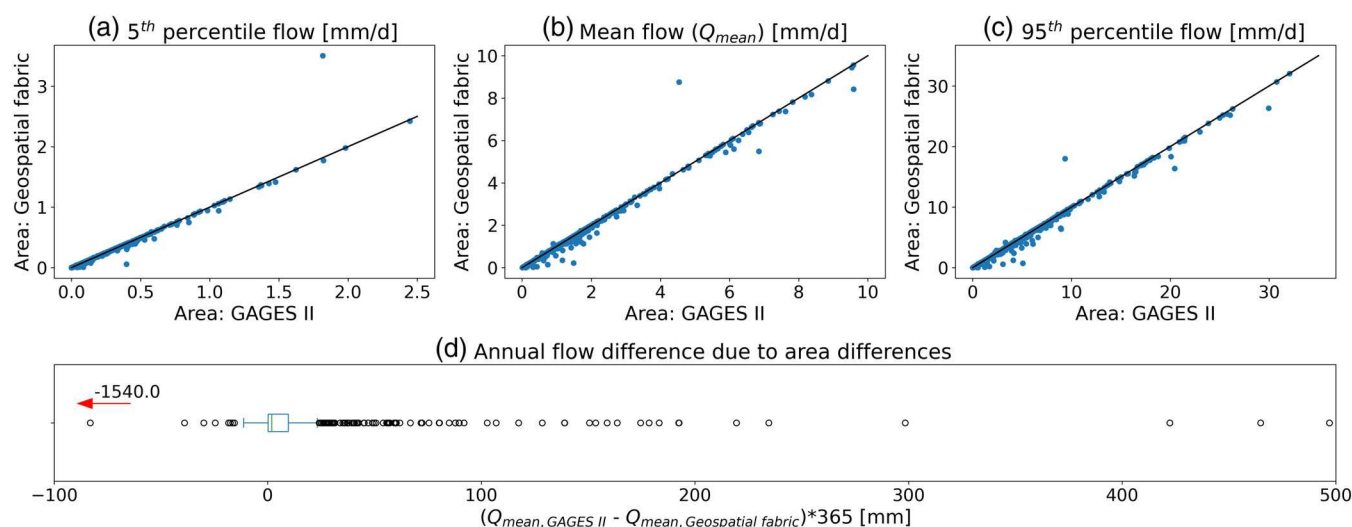
### 2.2.2 | High uncertainties in event rainfall and flow totals in Arizona

In the previously described investigation of 21 subcatchments of the San Pedro River (Southern Arizona), we used the TOSSH toolbox (Gnann, Coxon, et al., 2021) to calculate event rainfall and flow totals.

TOSSH plots rainfall totals against flow totals for a manual error check, and we found an unexpected lack of correspondence (example in Figure 4). We performed extended checks for other errors (such as errors in rainfall or flow series extraction, incorrect timestamps, or coding errors) but found none. We suggest that rainfall totals from the NLDAS-2 national gridded product do not accurately measure areal average rainfall for these small watersheds, due to a large grid size ( $0.125^\circ$ ), relatively sparse rain gauge network, and convective monsoon rainfall characterized by highly variable local accumulations. Further, the flow series contains significant periods of gap-filled data that may affect signature values: all three of the flow peaks in Figure 4 consist of gap-filled data. Our findings highlight the importance of considering local variations in data accuracy.

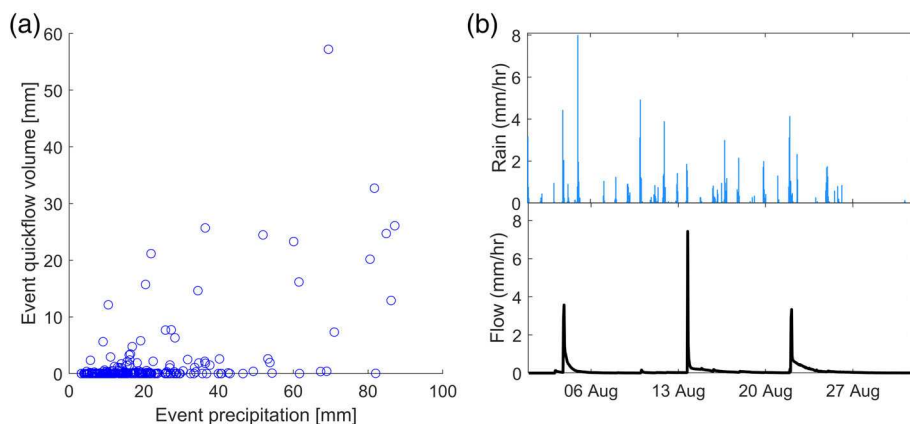
### 2.2.3 | Non-stationarity in soil moisture values invalidated magnitudes and dynamic range

Soil moisture data often exhibit increasing or decreasing trends over time. These trends may arise from non-stationarity in hydrologic processes or data quality issues, such as changes in sensor voltage power (Martini et al., 2015), oxidation of sensor rods, salinization, and soil compaction (Dorigo et al., 2013). These causes are difficult to distinguish and so trends are rarely addressed during quality control. Trends can impact various signatures, as estimated field capacity and wilting point signatures may change from year to year, and the shape of frequency distributions becomes unclear (Chandler et al., 2017; Araki et al., 2022; Figure 5). The dynamic range of soil moisture (the range between field capacity and wilting point, or between maximum and minimum values), often used as a normalization factor, may be overestimated. For some signatures, trending time-series should be excluded, such as for field capacity and wilting point estimates (Araki et al., 2022). Alternatively, observations can be detrended by subtracting a 2-year moving average signatures (Basak et al., 2017). An example of this approach is shown in Figure 5b, with the shape of



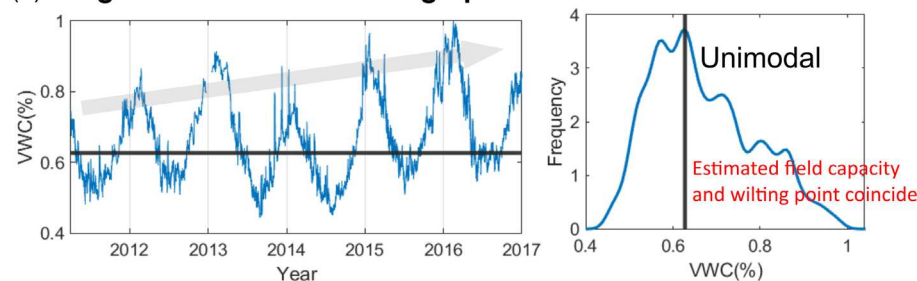
**FIGURE 3** Differences in magnitude-based signatures due to differences in catchment areas used to convert flows from [ $M^3T^{-1}$ ] to [ $MT^{-1}$ ]. (d) Red arrow shows the location of the final data point outside the axis limits.





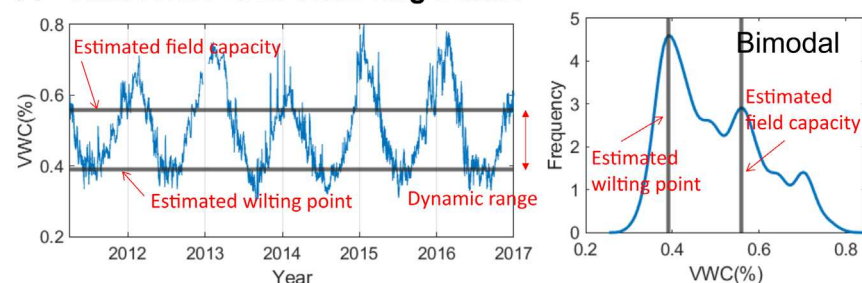
**FIGURE 4** Comparison of event totals for 2001–2022 and hourly time-series for August 2006 of areal-average NLDAS-2 rainfall and observed flow, USGS flow gauge 09470700 Banning Creek near Bisbee, watershed area 22.6 km<sup>2</sup>.

### (a) Original timeseries trending upward



**FIGURE 5** Example of the need to detrend a soil moisture time-series to enable correct calculation of distribution type and dynamic range signatures.

### (b) Timeseries after removing trends



frequency distributions and dynamic range calculated from the detrended time-series.

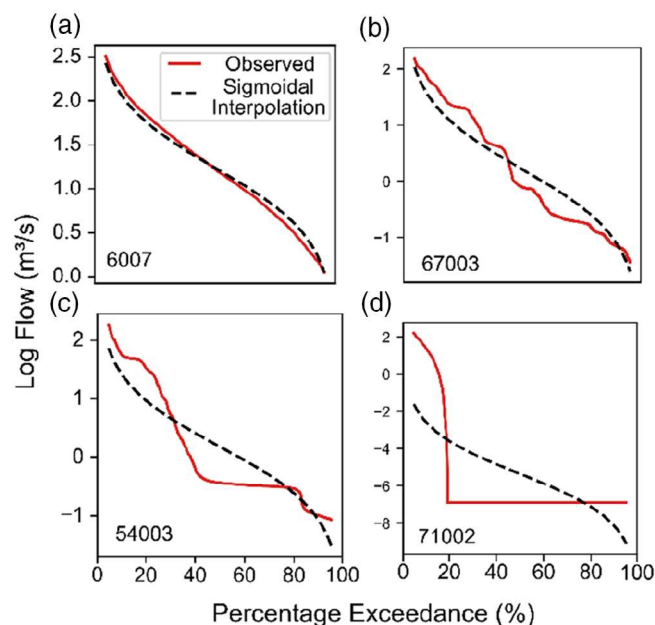
## 2.3 | Signature issues caused by human impacts

Many signatures for ecohydrology and hydrological processes were designed for natural flow regimes (McMillan, 2020; Yarnell et al., 2020). In human-impacted watersheds, signatures can fail if expected flow patterns no longer occur, and anthropogenic changes can override natural differences.

### 2.3.1 | Flow duration curve slope was modified for flow series impacted by water supply reservoirs

We used signatures to investigate how reservoirs impact flow regimes in Great Britain (Salwey et al., 2023). To maximize storage, operators

frequently release only the minimum flow required to protect downstream ecosystems (Maynard & Lane, 2012). Although pre-existing techniques (e.g., IHA) have been used to assess reservoir impact, we lacked the data to upscale these techniques (flow records do not pre-date reservoir construction, or insufficient locations with up-and-downstream gauges). Instead, we investigated signatures that require only a downstream flow series. We considered signatures that quantify flow duration curve (FDC) shape, because reservoirs may modify the full range of streamflow. While the most common FDC signature is the slope of its central portion (Yilmaz et al., 2008), this signature was unreliable in reservoir-impacted watersheds. Operations such as routine environmental flow releases cause flat segments and abrupt changes in the FDC, meaning that the central portion cannot be approximated by a linear slope (Figure 6a–d). Our solution was to develop a modified signature that quantifies deviations from an expected naturalized, sigmoidal FDC. High deviations reliably indicated large reservoirs. Despite this, we could not always distinguish natural and reservoir-impacted flow regimes. Natural



**FIGURE 6** Observed flow duration curves (red) and associated naturalized flow duration curves (black dashed) at example reservoir catchments in the U.K. (a) Ness at Ness-side - 6007; (b) Brenig at Llyn Brenig - 67 003; (c) Vyrnwy at Vyrnwy Reservoir - 54 003; (d) Hodder at Stocks Reservoir - 71 002).

groundwater gains can mimic pumped reservoir storage, and FDCs from ephemeral streams can mimic those where an upstream reservoir only releases water when full. Therefore, ignoring reservoirs in large sample studies risks mistaking reservoir-induced behaviours for unrelated natural processes.

### 2.3.2 | Stream temperature signatures are dominated by dam impacts

Human activity affects water quality signatures such as stream temperature indices that evaluate ecosystem health and nutrient cycling (Ficklin et al., 2023). Dams influence stream temperature regimes due to thermal stratification in reservoirs, with this water moved downstream via top release (warmer) or bottom release (colder) (Bonnema et al., 2020). For 57 eastern U.S. sites, we quantified thermal sensitivity, a signature that measures the strength of the relationship between air and water temperatures (Kelleher et al., 2012). In general, thermal sensitivity is lower in headwaters, where stream temperatures are buffered by groundwater and riparian shading, and higher in large watersheds where heat accumulates along the river. However, cool water releases below dams reduce thermal sensitivity values. Wade et al. (2023) trained random forest models on monthly thermal sensitivity values for 400 U.S. sites. They found that dam storage overwhelmed 23 other influences on river thermal regimes. Teasing out other influences required training models on sites with minimal dam impact. Dams also impact basic statistical metrics of water temperature: among 138 U.S. sites, dam-regulated sites had the warmest and

coolest 20-year trends in maximum monthly water temperatures (Kelleher et al., 2021). Overall, dam operations influence short-term behaviour and long-term trends in water temperature signatures.

## 3 | SUMMARY AND GUIDANCE FOR APPLYING SIGNATURES IN LARGE SAMPLE STUDIES

### 3.1 | Summary of case studies

In this commentary, we described multiple challenges that we encountered when calculating hydrologic signature values over large samples of watersheds (Table 1). Signatures designed for some hydroclimates may lead to incorrect interpretations in others, data quality may vary with location, and uncertainties may occur from unexpected sources. When signatures are used in human-impacted watersheds, careful analysis is required to separate human impacts from natural processes. When signatures are applied over continental or global datasets, all these factors should be considered when interpreting signature values and patterns.

### 3.2 | Guidance for signature users

We recommend that signature users be mindful of the range of precipitation or flow regimes for which signatures were designed. This design might have been intentional, for example, functional flow metrics that quantify ecologically-important flow features are designed for region-specific natural flow regimes (Yarnell et al., 2020) or inadvertent, for example, soil moisture signatures that were designed for NZ watersheds, all of which have similarities in climate (Branger & McMillan, 2020). Where a large sample analysis includes watersheds with different regimes, signatures might not perform correctly. This is particularly the case for human-impacted watersheds where reservoirs or other infrastructure may cause unusual flow patterns. Such patterns might cause signature codes to fail, or might mimic unrelated natural processes, for example, watersheds with groundwater gains can mimic those with pumped reservoir storage. Signatures may further be designed for a specific temporal resolution, for example signatures of flow alteration under hydropower require hourly data (Bevelhimer et al., 2015). We recommend that, especially where signature parameters must be selected, users should visually assess signature behaviour. For example, overlay event or recession periods on the hydrograph to check correct identification. For large samples, we encourage plotting for a subset of representative watersheds. Users should also monitor the structure of missing data, to avoid systematic biases. Biases should also be considered if sites or events are excluded from signature analyses as a solution to data quality or other issues (Table 1). Signature values can be compared with distributions of values from large sample studies (see signature distributions for the U.S., Great Britain, Australia and Brazil in McMillan et al. (2022)). Values outside or at the extremes of those distributions may indicate

**TABLE 1** Summary of case studies described in this commentary.

| Signatures affected                                                                  | Issue found                                                                                | Solution                                                                                                |
|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| <i>Precipitation regimes</i>                                                         |                                                                                            |                                                                                                         |
| Overland flow signatures in watersheds with daily convective rainfall                | Automatic event detection failed: merged or incorrectly associated events                  | Watershed-specific parameters, proposed improved signature algorithm allowing for time of concentration |
| Event runoff ratio in watersheds with rain on most days                              | Automatic event detection failed: incorrectly associated events                            | Manual checks and additional criteria for event rejection                                               |
| Soil moisture seasonality characterized by number of peaks in frequency distribution | Wider range of climates showed that bimodality has multiple causes beyond wet/dry periods  | Site-specific data inspection, proposed theoretical studies                                             |
| <i>Data Quality</i>                                                                  |                                                                                            |                                                                                                         |
| Flow magnitude signatures that normalize flow by watershed area                      | Watershed area is uncertain, two different estimates available                             | Reject watersheds with uncertain area, or include in uncertainty analysis                               |
| Signatures based on event totals for rainfall and runoff                             | Low correspondence of rainfall and runoff totals due to rainfall and flow uncertainty      | Exclude sites from analysis                                                                             |
| Soil moisture signatures for field capacity, wilting point and dynamic range         | Many sites have soil moisture trends caused by changes in processes or data quality issues | Exclude sites from analysis or detrend before use                                                       |
| <i>Human Impacts</i>                                                                 |                                                                                            |                                                                                                         |
| Mid-section slope of Flow Duration Curve                                             | Reservoir operations cause flat sections and abrupt slope changes                          | Assess reservoir impacts on FDC before applying standard signatures                                     |
| Thermal sensitivity and basic water temperature signatures                           | Effect of dam-controlled reservoirs override natural signals                               | Train models on sites without dam impacts                                                               |

errors, including those due to data quality. When interpreting high or low signature values, users should consider their perceptual model of watershed processes and how these affect the signature. We showed an example where bimodal soil moisture regimes have multiple interpretations; similarly, high BFI (base flow index) values could be caused by snow, permeable geology, or wetlands (Gnann, McMillan, et al., 2021). To facilitate such analyses, large sample perceptual model resources are becoming available (McMillan et al., 2023), although regional gaps still remain.

### 3.3 | Guidance for signature creators and coders

Many open-source toolboxes have been created to allow easier calculation of signatures. These include TOSSH (Gnann, Coxon, et al., 2021), eflows (Patterson et al., 2020), HydRun (Tang & Carey, 2017), HydroRecession (Arciniega-Esparza et al., 2017), HIA (Henriksen et al., 2006), Pastas (Collenteur et al., 2019) and others. We encourage toolbox authors to help reduce signature errors in large sample studies. One approach is to include warning flags to identify non-behavioural data series as part of signature code, such as a warning that few recession segments could be identified which may invalidate recession signatures. In addition, toolboxes can include plotting functions to allow rapid visual inspection of signature behaviour (as recommended above). Signature creators are in a good position to design relevant visual checks due to their expert knowledge of the signatures. Known issues, limitations and robustness of signatures can be shared in readme documents, such as is implemented for eflows documentation ([https://eflows.gitbook.io/project/known\\_issues](https://eflows.gitbook.io/project/known_issues)). Documents should specify requirements for the treatment of missing and gap-filled data. Through the combined efforts of signature creators and signature users, hydrologists can benefit from accurate signature calculations and interpretations in large sample studies.

#### ACKNOWLEDGEMENTS

Support to H. McMillan, G. Coxon, R. Araki and L. Bolotin was provided by the NSF Hydrologic Sciences Program, Division of Earth Sciences, Grant/Award Number 2124923.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

#### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

#### ORCID

Hilary McMillan  <https://orcid.org/0000-0002-9330-9730>

Gemma Coxon  <https://orcid.org/0000-0002-8837-460X>

Ryoko Araki  <https://orcid.org/0000-0002-3647-9768>

Saskia Salwey  <https://orcid.org/0000-0002-2100-3142>

Christa Kelleher  <https://orcid.org/0000-0003-3557-201X>

Yanchen Zheng  <https://orcid.org/0000-0002-2212-9134>

Wouter Knoben  <https://orcid.org/0000-0001-8301-3787>

Sebastian Gnann  <https://orcid.org/0000-0002-9797-5204>

Jan Seibert  <https://orcid.org/0000-0002-6314-2124>

Lauren Bolotin  <https://orcid.org/0000-0002-0295-9544>

#### REFERENCES

- Addor, N., Do, H. X., Alvarez-Garretón, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2020). Large sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, 65(5), 712–725.



- Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, 54(11), 8792–8812.
- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large sample studies. *Hydrology and Earth System Sciences*, 21, 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Almagro, A., Oliveira, P. T. S., & Brocca, L. (2021). Assessment of bottom-up satellite rainfall products on estimating river discharge and hydrologic signatures in Brazilian catchments. *Journal of Hydrology*, 603, 126897.
- Araki, R., Branger, F., Wiekenkamp, I., & McMillan, H. (2022). A signature-based approach to quantify soil moisture dynamics under contrasting land-uses. *Hydrological Processes*, 36(4), e14553. <https://doi.org/10.1002/hyp.14553>
- Araki, R., Mu, Y., & McMillan, H. (2023). Evaluation of GLDAS soil moisture seasonality in arid climates. *Hydrological Sciences Journal*, 68, 1109–1126. <https://doi.org/10.1080/02626667.2023.2206032>
- Arciniega-Esparza, S., Breña-Naranjo, J. A., Pedrozo-Acuña, A., & Appendini, C. M. (2017). Hydrorecession: A Matlab toolbox for streamflow recession analysis. *Computers & Geosciences*, 98, 87–92.
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C., Hasan, A., & Pineda, L. (2020). Global catchment modelling using world-wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, 24(2), 535–559.
- Basak, A., Mengshoel, O. J., Kulkarni, C., Schmidt, K., Shastri, P., & Rapeta, R. (2017). Optimizing the decomposition of time series using evolutionary algorithms: soil moisture analytics. Proceedings of the Genetic and Evolutionary Computation Conference GECCO '17, 1073–1080. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3071178.3071191>
- Beck, H. E., De Roo, A., & van Dijk, A. I. (2015). Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology*, 16(4), 1478–1501.
- Bevelhimer, M. S., McManamay, R. A., & O'Connor, B. (2015). Characterizing sub-daily flow regimes: Implications of hydrologic resolution on ecohydrology studies. *River Research and Applications*, 31(7), 867–879.
- Bonnema, M., Hossain, F., Nijssen, B., & Holtgrieve, G. (2020). Hydro-power's hidden transformation of rivers in the Mekong. *Environmental Research Letters*, 15(4), 044017. <https://doi.org/10.1088/1748-9326/ab763d>
- Botterill, T. E., & McMillan, H. K. (2022). Using machine learning to identify hydrologic signatures with an encoder-decoder framework. *Water Resources Research*, 59, e2022WR033091.
- Branger, F., & McMillan, H. K. (2020). Deriving hydrological signatures from soil moisture data. *Hydrological Processes*, 34(6), 1410–1427. <https://doi.org/10.1002/hyp.13645>
- Chandler, D. G., Seyfried, M. S., McNamara, J. P., & Hwang, K. (2017). Inference of soil hydrologic parameters from electronic soil moisture records. *Frontiers of Earth Science in China*, 5, 1–17. <https://doi.org/10.3389/feart.2017.00025>
- Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., & Schaars, F. (2019). Pastas: Open source software for the analysis of groundwater time series. *Groundwater*, 57(6), 877–885.
- Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J., Howden, N. J., Quinn, N., Wagener, T., & Woods, R. (2019). DECIPHeR v1: Dynamic fluxEs and connectivity for predictions of HydRology. *Geoscientific Model Development*, 12(6), 2285–2306.
- David, P. C., Chaffe, P. L., Chagas, V. B., Dal Molin, M., Oliveira, D. Y., Klein, A. H., & Fenicia, F. (2022). Correspondence between model structures and hydrological signatures: A large-sample case study using 508 Brazilian catchments. *Water Resources Research*, 58(3), e2021WR030619.
- Donnelly, C., Andersson, J. C., & Arheimer, B. (2016). Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, 61(2), 255–273.
- Dorigo, W. A. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D. D., Zamojski, D., et al. (2013). Global automated quality control of In situ soil moisture data from the international soil moisture network. *Vadose Zone Journal*, 12(3), vzj2012.0097. <https://doi.org/10.2136/vzj2012.0097>
- Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A., & Thompson, S. E. (2017). Event-scale power law recession analysis: Quantifying methodological uncertainty. *Hydrology and Earth System Sciences*, 21(1), 65–81.
- Ebeling, P., Kumar, R., Weber, M., Knoll, L., Fleckenstein, J. H., & Musolf, A. (2021). Archetypes and controls of riverine nutrient export across German catchments. *Water Resources Research*, 57(4), e2020WR028134.
- Ficklin, D. L., Hannah, D. M., Wanders, N., Dugdale, S. J., England, J., Klaus, J., Kelleher, C., Khamis, K., & Charlton, M. B. (2023). Rethinking river water temperature in a changing, human-dominated world. *Nature Water*, 1(2), 125–128. <https://doi.org/10.1038/s44221-023-00027-2>
- Giani, G., Tarasova, L., Woods, R. A., & Rico-Ramirez, M. A. (2022). An objective time-series-analysis method for rainfall-runoff event identification. *Water Resources Research*, 58, e2021WR031283. <https://doi.org/10.1029/2021WR031283>
- Gnann, S. J., Coxon, G., Woods, R. A., Howden, N. J., & McMillan, H. K. (2021). TOSSH: A toolbox for streamflow signatures in hydrology. *Environmental Modelling & Software*, 138, 104983.
- Gnann, S. J., McMillan, H. K., Woods, R. A., & Howden, N. J. (2021). Including regional knowledge improves baseflow signature predictions in large sample hydrology. *Water Resources Research*, 57(2), p.e2020WR028354.
- Grantham, T. E., Carlisle, D. M., Howard, J., Lane, B., Lusardi, R., Obester, A., Sandoval-Solis, S., Stanford, B., Stein, E. D., Taniguchi-Quan, K. T., & Yarnell, S. M. (2022). Modeling functional flows in California's Rivers. *Frontiers in Environmental Science*, 10, 787473.
- Henriksen, J. A., Heasley, J., Kennen, J. G., & Nieswand, S. (2006). *Users' manual for the Hydroecological integrity assessment process software (including the New Jersey assessment tools)*. U. S. Geological Survey.
- Janssen, J., & Ameli, A. A. (2021). A hydrologic functional approach for improving large-sample hydrology performance in poorly gauged regions. *Water Resources Research*, 57(9), p.e2021WR030263.
- Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, 11, 129–140.
- Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., & Marshall, L. (2012). Investigating controls on the thermal sensitivity of Pennsylvania streams. *Hydrological Processes*, 26(5), 771–785.
- Kelleher, C. A., Golden, H. E., & Archfield, S. A. (2021). Monthly river temperature trends across the US confound annual changes. *Environmental Research Letters*, 16(10), 104006. <https://doi.org/10.1088/1748-9326/ac2289>
- Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, 56, e2019WR025975. <https://doi.org/10.1029/2019WR025975>
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089–5110.
- Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., & Shalev, G. (2023). Caravan-a global community dataset for large sample hydrology. *Scientific Data*, 10(1), 61.
- Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6), 2863–2879.
- Magilligan, F. J., & Nislow, K. H. (2005). Changes in hydrologic regime by dams. 856. *Geomorphology*, 71(1–2), 61–78. <https://doi.org/10.1016/j.geomorph.2004.08.017>

- Martini, E., Wollschläger, U., Kögler, S., Behrens, T., Dietrich, P., Reinstorf, F., Schmidt, K., Weiler, M., Werban, U., & Zacharias, S. (2015). Spatial and temporal dynamics of hillslope-scale soil moisture patterns: Characteristic states and transition mechanisms. *Vadose Zone Journal*, 14(4), vzj2014.10.0150. <https://doi.org/10.2136/vzj2014.10.0150>
- Massmann, C. (2020). Identification of factors influencing hydrologic model performance using a top-down approach in a large number of US catchments. *Hydrological Processes*, 34(1), 4–20.
- Maynard, C. M., & Lane, S. N. (2012). Reservoir compensation releases: Impact on the 863 macroinvertebrate Community of the Derwent River, Northumberland, U.K.-a 864 longitudinal study. *River Research and Applications*, 28(6), 692–702. <https://doi.org/10.1002/rra.2557>
- McMillan, H. (2020). Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, 34(6), 1393–1409.
- McMillan, H., Araki, R., Gnann, S., Woods, R., & Wagener, T. (2023). How do hydrologists perceive watersheds? A survey and analysis of perceptual model figures for experimental watersheds. *Hydrological Processes*, 37(3), e14845.
- McMillan, H. K. (2021). A review of hydrologic signatures and their applications. *Wiley Interdisciplinary Reviews: Water*, 8(1), e1499.
- McMillan, H. K., Booker, D. J., & Cattoën, C. (2016). Validation of a national hydrological model. *Journal of Hydrology*, 541, 800–815.
- McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., & Woods, R. A. (2011). Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. *Hydrological Processes*, 25(4), 511–522.
- McMillan, H. K., Gnann, S. J., & Araki, R. (2022). Large scale evaluation of relationships between hydrologic signatures and processes. *Water Resources Research*, 58(6), e2021WR031751.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., & Duan, Q. (2015). Development of a large sample watershed-scale hydrometeorological dataset for the contiguous USA: Dataset characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19, 209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Patterson, N. K., Lane, B. A., Sandoval-Solis, S., Pasternack, G. B., Yarnell, S. M., & Qiu, Y. (2020). A hydrologic feature detection algorithm to quantify seasonal components of flow regimes. *Journal of Hydrology*, 585, 124787.
- Peek, R., Irving, K., Yarnell, S. M., Lusardi, R., Stein, E. D., & Mazor, R. (2022). Identifying functional flow linkages between stream alteration and biological stream condition indices across California. *Frontiers in Environmental Science*, 9, 638.
- Salwey, S., Coxon, G., Pianosi, F., Singer, M. B., & Hutton, C. (2023). National-Scale Detection of reservoir impacts through hydrological signatures. *Water Resources Research*, 59, e2022WR033893.
- Sehgal, V., & Mohanty, B. (2023). Preferential hydrologic states and tipping characteristics of global surface soil moisture. *ESS Open Archive*, 1–39. <https://doi.org/10.22541/essoar.167840001.13313960/v1>
- Smith, B. K., & Smith, J. A. (2015). The flashiest watersheds in the contiguous United States. *Journal of Hydrometeorology*, 16(6), 2365–2381.
- Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., & Tallaksen, L. M. (2020). Beyond binary baseflow separation: A delayed-flow index for multiple streamflow contributions. *Hydrology and Earth System Sciences*, 24(2), 849–867.
- Tang, W., & Carey, S. K. (2017). HydRun: A MATLAB toolbox for rainfall-runoff analysis. *Hydrological Processes*, 31(15), 2670–2682.
- Tarasova, L., Basso, S., Zink, M., & Merz, R. (2018). Exploring controls on rainfall-runoff events: 1. Time series-based event separation and temporal dynamics of event runoff response in Germany. *Water Resources Research*, 54, 7711–7732. <https://doi.org/10.1029/2018WR022587>
- Tashie, A., Pavelsky, T., & Band, L. E. (2020). An empirical reevaluation of streamflow recession analysis at the continental scale. *Water Resources Research*, 56(1), e2019WR025448.
- Tyralis, H., Papacharalampous, G., Langousis, A., & Papalexiou, S. M. (2021). Explanation and probabilistic prediction of hydrological signatures with statistical boosting algorithms. *Remote Sensing*, 13(3), 333. <https://doi.org/10.3390/rs13030333>
- Wade, J., Kelleher, C., & Hannah, D. M. (2023). Machine learning unravels controls on river water temperature regime dynamics. *Journal of Hydrology*, 623, 129821. <https://doi.org/10.1016/j.jhydrol.2023.129821>
- Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, 19(9), 3951–3968.
- Yarnell, S. M., Stein, E. D., Webb, J. A., Grantham, T., Lusardi, R. A., Zimmerman, J., Peek, R. A., Lane, B. A., Howard, J., & Sandoval-Solis, S. (2020). A functional flows approach to selecting ecologically relevant flow metrics for environmental flow applications. *River Research and Applications*, 36(2), 318–324.
- Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, 44(9), W09417. <https://doi.org/10.1029/2007WR006716>
- Zheng, Y., Coxon, G., Woods, R., Li, J., & Feng, P. (2023a). Controls on the spatial and temporal patterns of rainfall-runoff event characteristics – a large sample of catchments across Great Britain. *Water Resources Research*, 59, e2022WR033226. <https://doi.org/10.1029/2022WR033226>
- Zheng, Y., Coxon, G., Woods, R., Li, J., & Feng, P. (2023b). A framework for estimating the probability distribution of event runoff coefficient in ungauged catchments. *Water Resources Research*, 59, e2022WR033227. <https://doi.org/10.1029/2022WR033227>
- Zmijewski, N., & Wörman, A. (2016). Hydrograph variances over different timescales in hydropower production networks. *Water Resources Research*, 52, 5829–5846. <https://doi.org/10.1002/2015WR017775>

**How to cite this article:** McMillan, H., Coxon, G., Araki, R., Salwey, S., Kelleher, C., Zheng, Y., Knoben, W., Gnann, S., Seibert, J., & Bolotin, L. (2023). When good signatures go bad: Applying hydrologic signatures in large sample studies. *Hydrological Processes*, 37(9), e14987. <https://doi.org/10.1002/hyp.14987>