Quantification and Analysis of Hydrograph Behavior Using Groundwater Signatures

Collenteur, Raoul A.; Vonk, Martin A.; Haaf, Ezra

DOI
10.1111/gwat.13486

Publication date
2025

Document Version
Final published version

Published in
Groundwater

Groundwater

Methods Brief/

Quantification and Analysis of Hydrograph Behavior Using Groundwater Signatures

by Raoul A. Collenteur[1], Martin A. Vonk[2,3], and Ezra Haaf[4]

Abstract

The study of hydraulic head changes over time is a common task for groundwater hydrologists. Groundwater signatures are numerical metrics, or statistical aggregates, that quantify the behavior observed in hydraulic head hydrographs. Signatures can be helpful in a number of classical hydrological tasks, such as hydrograph classification, clustering, change detection, and model evaluation, selection, and calibration. Despite the potential benefits of using signatures in groundwater studies, their application has not yet been thoroughly explored. To support research into the application of signatures in groundwater studies, we introduce the new groundwater signatures module from the Pastas software. The signatures module is written in Python, fully tested and documented, and available as open-source software under the MIT license. In this paper, it is shown how the signatures are tested and can be used in practical applications through two examples. In the first example, signatures are used to characterize and cluster monitoring wells in a nationwide monitoring network in Switzerland. In the second example, signatures are used to evaluate how well different groundwater model structures simulate the heads. Future research opportunities involving groundwater signatures are discussed.

Introduction

Groundwater signatures are numerical metrics or statistical aggregates that quantify the behavior observed in hydraulic head hydrographs. These signatures can range from simple statistics, such as the mean and range, to more complex metrics such as the duration of low pulses, the seasonality, and the rise rate after recharge events. The use of hydrological signatures (also referred to as features, indices, or [fingerprint] metrics) is common in other parts of the hydrological sciences (e.g., streamflow, soil moisture, and eco-hydrology), with applications ranging from process characterization and classification to model calibration and evaluation (see McMillan 2021, and references therein). Recently, the signature concept has been translated and further developed for applications in groundwater studies, among others, by Heudorfer et al. (2019), Haaf et al. (2020), and Giese et al. (2020).

We examine two time series of hydraulic heads to illustrate how signatures are used to quantify groundwater system behavior. The time series are shown in Figure 1, along with the cumulative frequency distributions and the values of six groundwater signatures. Visually, these time series are rather different, with the upper one showing more flashy behavior compared to the lower one, which shows a smoother hydrograph. The signature values in the table in Figure 1 show clear differences that match this visual interpretation. For example, the heads in the upper hydrograph show a higher rise rate (the average increase in head in $\mathrm{cm\,d^{-1}}$) and a lower base level stability (a measure of how variable the head is compared to a moving-average). This indicates a faster responding,

[1]Department Water Resources and Drinking Water, Eawag, Dübendorf, Zürich, Switzerland
[2]Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, South Holland, The Netherlands
[3]Artesia B.V., Schoonhoven, South Holland, The Netherlands
[4]Corresponding author: Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden; ezra.haaf@chalmers.se

Article impact statement: Groundwater signatures are a powerful tool for the characterization of hydrograph behavior, implemented in a module in the Python package Pastas.

Received October 2024, accepted March 2025.

doi: 10.1111/gwat.13486

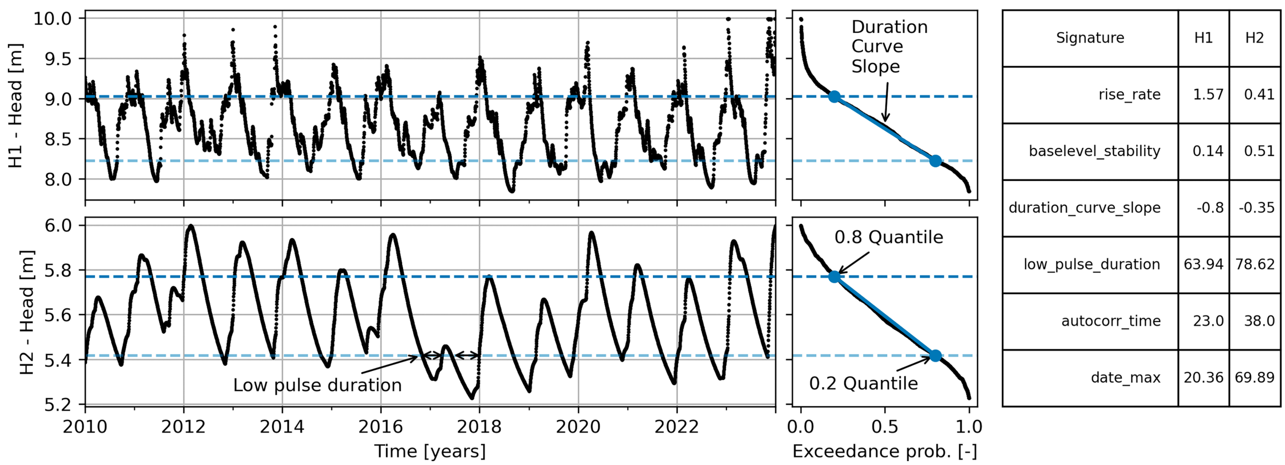

Figure 1. Example of two head time series, the related cumulative frequency distributions, and different groundwater signatures. Note how the signature values differ between the head time series.

flashier system, compared to the lower hydrograph. This also results in a higher variability of the heads in the upper plot, as quantified by the steeper negative slope of the duration curve. The autocorrelation time, the number of days it takes the autocorrelation function to drop below a threshold (here, 0.8), and the duration of low pulses (head events below the 0.8 quantile) are both larger for the lower hydrograph, indicating a longer system memory. Finally, the average day of the year when the maximum occurs (date_max) is later for the lower hydrograph, suggesting a more buffered system (both wells are in the same area).

In the example above, groundwater signatures were used to compare two head time series and characterize the groundwater systems (i.e., fast vs. slow). Research so far has primarily focused on the use of signatures for characterization and classification (e.g., Haaf and Barthel 2018; Heudorfer et al. 2019; Rinderer et al. 2019; Nolte et al. 2024), or identification of reference wells (e.g., Wunsch et al. 2022). Haaf et al. (2020) and Giese et al. (2020) showed that it is possible to relate groundwater signatures to physical processes or environmental and climatic descriptors. Such information can increase our understanding of the underlying causes of observed groundwater dynamics, and ultimately be used to predict groundwater dynamics.

Signatures may also be useful in other common tasks of groundwater hydrologists besides those outlined above, similar to how they are used in other parts of hydrology. Perhaps most notably, signatures may be helpful in various stages of groundwater modeling, that is, calibration, evaluation, and selection. Open-source software, well-documented and tested, would help a community effort to explore such applications. Therefore, this paper introduces the signatures module, part of the Python package Pastas (Collenteur et al. 2019a), to compute signatures of time series of hydraulic heads. The objectives of this paper are to (1) familiarize readers with the groundwater signature concept and the Pastas signatures module, and (2) encourage the exploration and application of signature-like concepts in groundwater studies through example applications and the discussion of potential use cases.

The remainder of this paper is structured as follows. In the following section, the Pastas signatures module is introduced and tested using synthetic head data generated with a numerical variably saturated groundwater model (Vonk et al. 2024). In two example applications, the use of signatures is exemplified. In the first example, signatures are used to characterize and cluster monitoring wells in a nationwide monitoring network in Switzerland. In the second example, signatures are used to evaluate how well different groundwater model structures simulate the heads. The paper concludes with a discussion on the potential applications of groundwater signatures and concluding remarks.

The Pastas Signatures Module

Module Description

The signatures module is developed as part of the open-source Python package Pastas to model and analyze groundwater time series (Collenteur et al. 2019a). The module can be used independently of other Pastas modules and is accessible from the `pastas.stats` sub-package. The module is written in pure Python code, hosted on GitHub (www.github.com/pastas/pastas, last accessed February 19, 2025). All signature methods are fully documented, with references to the literature describing the original signatures. An example notebook is available to show how to compute the signatures on the documentation website (pastas.readthedocs.io, last accessed February 19, 2025).

The Pastas signatures module contains 31 signatures at the time of publication. Table 1 provides an overview of the available signatures, along with a brief description of the signature and its units. Many of the signatures are based on the work of Heudorfer et al. (2019) and the references therein, but were substantially modified to pass a comprehensive set of tests discussed below. For the complete descriptions, formulas, and references for the individual signatures, we refer to the documentation website and the source code of the individual signatures.

Basic Usage

The Pastas signatures module makes use of the open-source Python package Pandas (McKinney et al. 2010). Pandas is a popular tool for data analysis and can easily read tabular data files (e.g., CSV-files) and handle time series data. All signature methods take a `Pandas.Series` with a `DatetimeIndex` as input. The computation of individual signatures, for example, the average duration of the low pulses, is done as follows:

```
import pandas as pd

from pastas.stats import signatures



head = pd.read_csv(

    "heads.csv",

    index_col="datetime",

    parse_dates=["datetime"],

).squeeze()

sig = signatures.low_pulse_duration(head)
```

which returns a single signature value (a float). An advantage of this approach is that one can carefully control the input arguments to the individual methods. For example, one may adapt the lower quantile used to determine when a low pulse starts (e.g., from 0.2 to 0.1 to select more rare events). For the possible input arguments, which differ per method, the reader is referred to the documentation of the individual methods.

Often there is a need to compute a comprehensive set of signatures for one or more head time series. This can be done using the `summary` method as follows:

```
sigs = signatures.summary(head)
```

where the return variable `sigs` is a `Pandas.DataFrame` with the signature names as index, and the signatures values as columns (number of columns depending on the number of head time series provided). A single time series may be provided as a `Pandas.Series`, and a set of time series as a `Pandas.DataFrame`, both with a `DatetimeIndex`. By default, this method will compute all the available signatures, but it is possible to compute a specified set of signatures by providing a list with signature names.

Testing the Signatures

As noted by McMillan et al. (2023), it is important to test signatures for good behavior before usage. This is particularly true when designing or developing software meant for wider usage, as is done here. In McMillan et al. (2017), five guidelines for the design and choice of (a set of) hydrological signature(s) are suggested. Four of these, discriminatory power, identifiability, robustness, and consistency, are adapted to groundwater signatures and used here. The fifth guideline, representativeness, relates to the spatial scale of the signatures. It requires that the signatures reflect the average behavior across a catchment or, in the case of groundwater, across an aquifer. However, the head in a single well generally does not represent the behavior of an entire aquifer and is by definition sensitive to its location in the aquifer. Investigating if signatures can be used to assess how representative a well is of an aquifer is outside the scope of this study. Therefore, the concept behind this guideline was not applied in the analysis. The four applied guidelines are as follows:

1. *Discriminatory power*: The signature should be able to distinguish between different groundwater systems that are identified by a hydrologist as different.
2. *Identifiability*: The uncertainty of the signature value due to uncertainty in the head data should be small compared to the range of the signature value.
3. *Robustness*: The value of the signature should be robust to common characteristics of the time series (e.g., different measurement frequencies and periods). If the measurement frequency, period, or length is changed, this should not substantially change the signature value.
4. *Consistency*: The signature values should be comparable between different aquifer systems. This is often hampered by large differences in drainage bases between aquifers, causing large absolute differences in the heads.

All signatures were (re-)designed, implemented, and thoroughly tested with the above four guidelines in mind. Below, we discuss how the signatures were tested for the first three guidelines. The fourth guideline, consistency, can be dealt with through normalization of the heads (when applicable) to values between zero and one by subtracting the minimum and dividing over the range as suggested in Heudorfer et al. (2019). Table 1 shows which signatures are normalized by default or not.

Testing if the signatures behave in accordance with the four guidelines using real-world data is challenging. The primary causes are that it is unknown if a groundwater system changes through time and if two systems are really different. The signatures were therefore tested on synthetic hydraulic head data generated with a variably saturated groundwater model developed in Vonk et al. (2024). This model simulates hydraulic heads in an aquifer between two parallel canals, with recharge from above. Simulations vary with three aquifer soil types: sand, sandy loam, and silt loam, and nine constant river stages, causing the unsaturated zone thicknesses to vary between zero and approximately five meters. Five virtual boreholes, evenly distributed from the middle of the aquifer toward the river, are implemented to obtain the synthetic heads.

Table 1
Signature Methods Currently Available From the Pastas Signature Module With a Description and the Units

Method Name	Description	Units	Norm.
cv_period_mean	Coefficient of variation (CV) of the mean heads over a period (default monthly)	—	
date_min/date_max	Average day of the year when the minimum/maximum head occurs, computed using circular statistics	T	
cv_date_min/cv_date_max	Coefficient of variation of the day of the year when the minimum/maximum head occurs, computed using circular statistics	—	
rise_rate/fall_rate	Mean of the positive/negative head changes from 1 d to the next	L/T	
cv_fall_rate/cv_rise_rate	Coefficient of variation of the fall/rise rate	—	
parde_seasonality	Difference between the maximum and minimum Pardé coefficient, which is the ratio of the mean monthly head to the mean annual head	—	✓
avg_seasonal_fluctuation	Mean annual difference between the averaged 3 highest monthly heads per year and the averaged 3 lowest monthly heads per year	L	
interannual_variation	The average of the range in annually averaged 3 highest and 3 lowest monthly heads	L	
high/low_pulse_count	Average number of times the head is above/below a certain quantile threshold per year (default is 0.2 for low pulses and 0.8 for high pulses)	—	
high/low_pulse_duration	Average duration of pulses where the head is above/below a certain quantile threshold per year (default is 0.2 for low pulses and 0.8 for high pulses)	T	
bimodality_coefficient	Squared product moment skewness plus one, divided by product moment kurtosis	—	✓
mean_annual_maximum	Mean of the annual maximum of the normalized heads	—	✓
reversals_avg	Average number of changes in direction (fall or rise) in the daily heads per year	—	
reversals_cv	Coefficient of variation of the annual number of changes in direction in the daily heads	—	
colwell_contingency	Colwell's contingency, a measure for repeatability of seasonal patterns	—	✓
colwell_constancy	Colwell's constancy, a measure how uniformly events happen throughout all seasons	—	✓
recession_constant	Value of the exponential decay constant in days, fitted on all recession segments	T	
recovery_constant	Value of the exponential constant in days, fitted on all recovery segments	T	
duration_curve_slope	Slope of the head duration curve between a lower quantile (default 0.1) and an upper quantile (default 0.9)	1/L	
duration_curve_ratio	Ratio of the head duration curve between the lower and upper quantiles (default 0.1 and 0.9)	—	✓
richards_path length	The path length of the time series, standardized by time series length and median	—	✓
baselevel_index	The total sum of the heads divided by the total sum of the base level head. The base level head is defined as the minimum head over a 30-day moving window	—	✓
baselevel_stability	The difference of maximum and minimum annual base level.	L	✓
magnitude	Difference between the minimum and maximum heads, divided by the minimum head	—	
autocorr_times	Number of days it takes the autocorrelation to drop below a threshold (default is 0.8)	T	

Note: The right column shows whether the heads are normalized before computing the signature or not.

In total, the three soil types, nine river stages, and five borehole locations result in 135 daily head time series over 30 years (1993 through 2023). It is noted here that for our purpose, we re-ran the model to obtain head time series for a period of 30 years, rather than the 22 years of the original work. For further details on the model used to generate this synthetic head data set, we refer to Vonk et al. (2024).

Discriminatory Power

To assess the discriminatory power of the signatures, the signatures were calculated for three of the synthetic head time series from Vonk et al. (2024), generated using three different aquifer soil types (sand, silt-loam, and sand-loam, see Figure 2a). These were identified as hydrologically different by the Authors, and the signatures are expected to be able to discriminate between them,

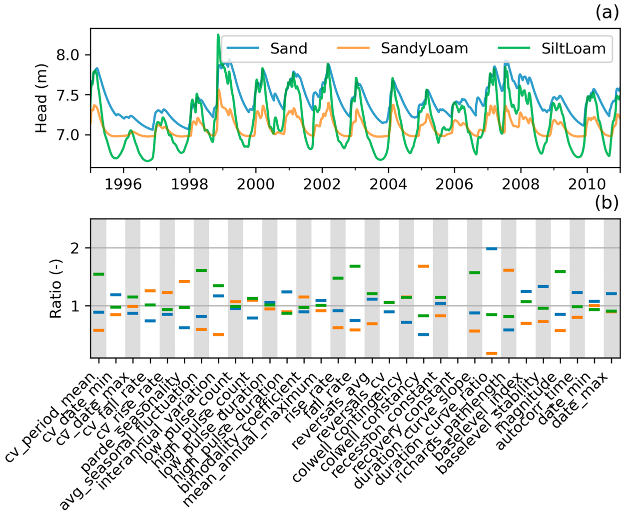

Figure 2. Three selected time series of hydraulic heads (a) and the related normalized groundwater signatures (b). To test the discriminatory power of the signatures, the signatures were normalized by dividing by the mean of each signature for the three heads. The virtual boreholes of the selected time series are located at 1/4 of the distance between the parallel rivers (see Vonk et al. 2024, for more details).

even though the time series are a result of the same meteorological time series. The plot in Figure 2b shows the signature values for the three head time series, normalized by the mean of the three signature values per signature. This analysis shows that most of the signatures have different values for these three time series. The signature values range between 10% and 250% of the mean signature, which indicates that signature values for these three time series are diverse and can potentially discriminate between these three groundwater systems.

Testing Identifiability and Robustness

To test the identifiability, the signatures computed on the original synthetic head time series were compared to the signatures computed on noisy head time series. To construct noisy head time series, normally distributed errors with a mean of zero and a standard deviation equal to 5% of the standard deviation of the heads were added to the heads.

The robustness of the signatures was tested for three common characteristics of hydraulic head time series: different measurement frequencies, periods, and time series lengths, as follows:

1. The impact of the measurement frequency was tested by computing the signatures on all 135 time series of 30 years sampled with three commonly used measurement frequencies: daily, weekly, and monthly values. Signatures calculated on weekly and monthly data were normalized by dividing through the signatures computed on daily values. The daily signatures were thus used as a benchmark here.
2. The impact of using different measurement periods was tested by computing the signatures on a moving 10-year window from 1993 to 2023, for three selected head time series (shown in Figure 2a). The signature values were normalized by the average value of each signature over all periods for each time series.
3. The impact of the time series length was tested by computing the signature values on different slices of all the 135 time series with different lengths (2, 5, and 10 calendar years). The values were normalized by the signature values computed over the entire 30-year time series length.

Detailed results from these analyses can be found in the Supporting Information. Here we provide a summary of the results, with the aim of aiding the selection of signatures for a particular data set or use case. A highlight table summarizing how well signatures adhere to the principles of identifiability and robustness is provided in Figure 3. The values in the table show the median absolute percentage deviation calculated from the “true” to modified signature, for each test case, as described above. It is noted here that the wording “true” signature value is only used here to indicate the value that would be computed for the entire or unmodified time series (here 30 years). The larger the deviation (yellow and red colors), the less identifiable or robust the signature is to the tested time series characteristic.

As depicted in Figure 3, the majority of signatures exhibit high identifiability both when the hydrographs are smoothed or noisy. There are exceptions, however, when the signatures computed on noisy or smoothed data substantially differ from the ‘true’ signatures. These generally are signatures calculated based on the characteristics of the groundwater response to individual stress pulses, such as recharge events. Examples include the low and high pulse counts, the average number of reversals, and the recession and recovery constant. These signatures should thus be avoided when working with noisy data. Smoothing the data may help improve the estimation of some signatures, for example the low and high pulse counts and durations.

The robustness of signatures to changes in measurement frequencies, the time periods used for their computation, and the time series length is more intricate. The analysis suggests that all signatures can be computed on a weekly basis instead of daily, with minimal impact on the signature value. Using monthly values, however, will substantially affect approximately one-third of the signatures (here defined as median deviation exceeding 10%). As a guideline, we recommend that these sensitive signatures should only be used to compare time series with approximately monthly values exclusively, or daily and weekly values. Using either of these frequencies should work, but mixing them should be avoided.

About a third of the signatures are substantially influenced by the choice of the period used for computation (i.e., 1995–2005 or 2010–2020). This factor should be considered when analyzing signatures extracted from non-concurrent time series. The length of the time series also impacts the signatures values (i.e., 2, 5, or 10 years), compared to using the entire 30-year period. Clearly,

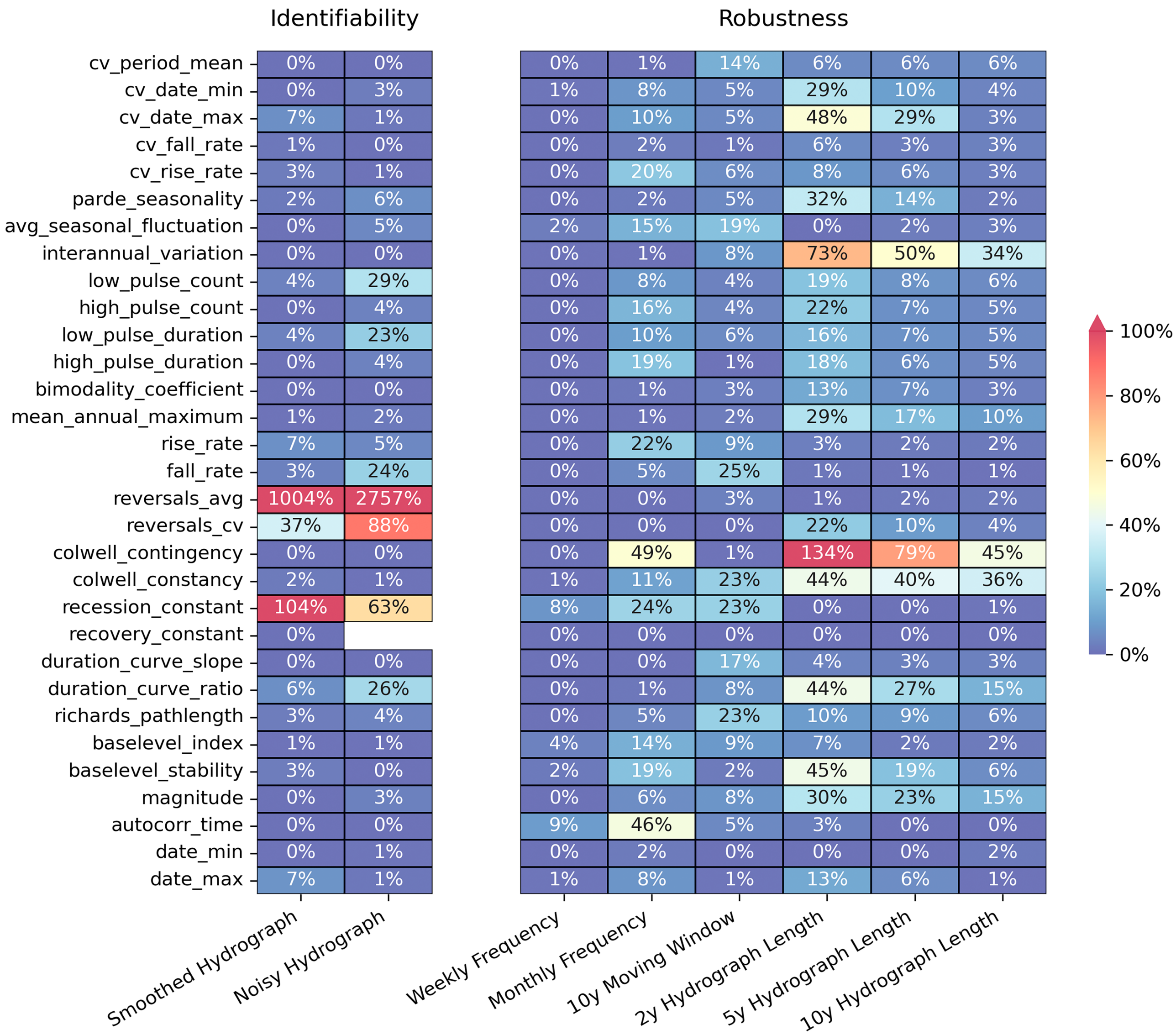

Figure 3. Highlight table of median signature test scores in terms of percentage deviation from "true" signature value (i.e., the unmodified or full 30-year time series). The larger the deviation (yellow and red colors), the less identifiable or robust the signature is to the tested time series characteristic. The color scale is capped off at 100% deviation.

the signature values get closer to those computed on the entire 30-year period when using longer time series. Large deviations are still observed for most signatures for 2-and 5-year time series lengths, but substantially smaller differences are generally found when using 10 years of data. These results add nuance to the findings of Heudorfer et al. (2019) who found that, based on shorter (20 years) but measured time series, the signatures tend to stabilize after 8 years. Based on the results of the tests using synthetic data and the results from Heudorfer et al. (2019) we recommend computing the signatures on approximately 10 years of data or longer, as a general rule of thumb.

Example Applications

Below, two examples are given of how signatures may be used to help solve common hydrological tasks. The first task is to characterize and cluster monitoring wells into separate clusters based on their hydrological behavior. The second task is to evaluate the performance of different groundwater models, each representing a different conceptualization of the groundwater system.

Characterization and Clustering

In this example, it is shown how to characterize and cluster monitoring wells using groundwater signatures. Data from the Swiss national groundwater monitoring network (NAQUA), operated by the Swiss Federal Office for the Environment (FOEN), are used as an example. One advantage of using signatures instead of the raw time series for this task is that time series with different measurement periods can also be clustered on their behavior (see Figure 2, Robustness, 10y Moving Window). The data set contains head time series from 28 monitoring wells throughout Switzerland, covering different climates and aquifer types. A more detailed description of the dataset and the hydrogeological settings is provided in Collenteur et al. (2023).

The clustering is performed as follows. First, all available signatures are computed for the data. Second, the signatures values are normalized to values between zero and one by subtracting the minimum and dividing by the value range. This ensures that the signatures have equal weights in the clustering algorithm. Third, Spearman correlations between the signatures are computed to identify highly correlated and thus non-unique signatures. These correlations are used to select a set of signatures with correlations less than 0.8 (an arbitrary threshold chosen for illustrative purposes). In the fourth step, the agglomerative hierarchical clustering (AHC) approach using Ward's method (available from Scikit-learn, Pedregosa et al. 2011) is applied for the clustering. This is a common approach for clustering based on signatures (e.g., Haaf and Barthel 2018; Nolte et al. 2024), but it is noted here that other algorithms or approaches (e.g., Wunsch et al. 2022) can be used as well. The monitoring wells were clustered into five clusters of different sizes. The final number of clusters was determined by testing different numbers of clusters and visually analyzing the result.

Figure 4 shows the results from the clustering exercise. The plot on the left-hand side shows a tree diagram (or dendrogram) of the different clusters, and the right-hand side shows the normalized head time series for each cluster. We emphasize that the clustering was done on the signatures and not on the head time series. The time series in the clusters should thus show similar behavior, not necessarily temporal resemblance. The clusters generated by the AHC approach indeed appear to show similar behavior. For example, cluster C0 shows slowly responding heads, while cluster C3 shows heads that respond faster. Cluster C4 contains one monitoring well (Wila) with capped off heads. This unique time series in the data set has a high bimodality coefficient. The approach outlined above can be used to cluster large data sets of hydraulic head time series with different measurement frequencies and periods.

Groundwater Model Evaluation

In the example above, signatures were used only to quantify the behavior observed in the measured heads. Signatures are also helpful in conjunction with groundwater models. In this example, signatures are used to evaluate three models used to simulate hydraulic heads. This form of model evaluation is already common in streamflow modeling (see, e.g., Euser et al. 2013). The core idea is to evaluate how well the signatures computed on the simulated heads mimic the signatures computed on the measured heads. To illustrate this use case, we use three different lumped parameter groundwater models from the Pastas package (Collenteur et al. 2019a) to simulate one of the synthetic heads from Vonk et al. (2024). It is noted, however, that signature evaluation can also be used for other types of groundwater models.

The basic workflow for the evaluation of groundwater models using signatures is as follows. First, the goal of the modeling is defined. Here, we aim to better understand which model structure better simulates the groundwater behavior important for groundwater droughts, and in turn, which processes should be included in the model. Second, the heads were simulated using three alternative models (M1, M2, and M3) from the Pastas software. The first model (M1) computes recharge as a linear combination of precipitation and potential evaporation. The other two models (M2 and M3) use a soil-water balance approach to compute nonlinear recharge from these two fluxes. The difference between M2 and M3 is that M2 does not allow

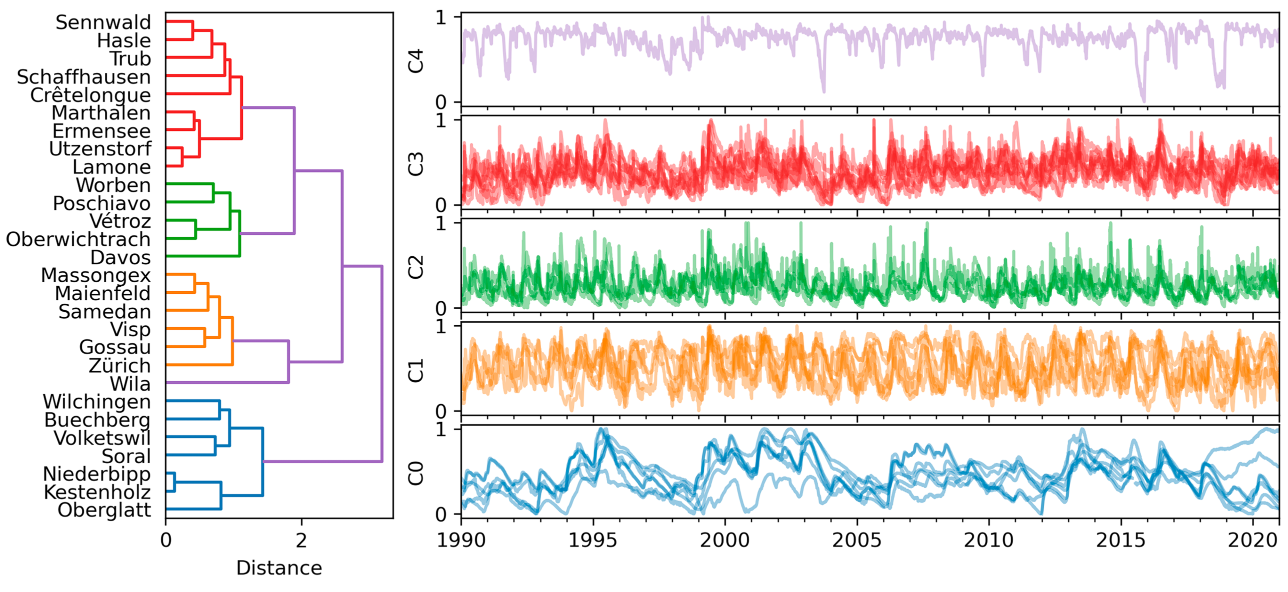

Figure 4. Results of the clustering exercise using groundwater signatures into five clusters of hydraulic head time series. The left-hand plot shows the dendrogram of the clustering, and the right-hand plot shows the five clusters of time series as identified here.

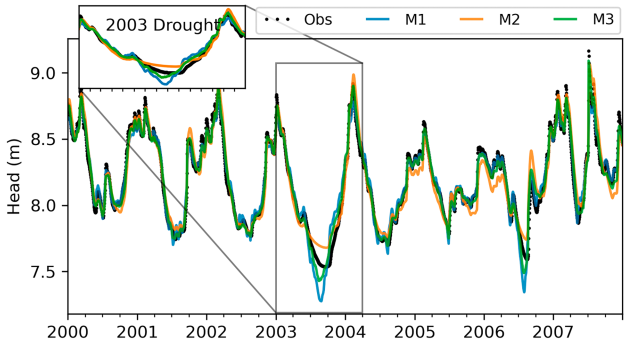

Figure 5. Measured and simulated heads with three different lumped parameter models (M1, M2, and M3). The inset plot shows the head during the 2003 drought event.

TABLE 2
Evaluation Criteria for the Different Signatures (E_{sig} for Each Model, the Euclidean Distance Metric, and Nash-Sutcliffe Efficiency (NSE)

	M1	M2	M3
$E_{\text{rise\_rate}}$	0.81	**0.87**	0.83
$E_{\text{fall\_rate}}$	**0.99**	0.77	0.85
$E_{\text{autocorr\_time}}$	0.92	0.96	**1.00**
$E_{\text{baselevel\_stability}}$	0.48	**0.97**	0.91
$E_{\text{low\_pulse\_duration}}$	0.80	**0.86**	0.77
$E_{\text{date\_min}}$	0.93	**1.00**	0.91
E_d	0.84	**0.91**	0.88
NSE	0.96	0.93	**0.98**

Note: The best score for each row is highlighted with a bold font.

uptake of groundwater for evaporation, while M3 does include this process in a simplified way, by subtracting the leftover potential evaporation from the recharge flux. The measured and simulated heads are shown in Figure 5.

The third step is to select the signatures to be used and to compute the signatures on the measured heads ($S_n(h_{\text{obs}})$) and the simulated heads ($S_n(h_{\text{sim}})$). Here, six signatures are selected that quantify behavior thought to be related to drought events: the rise rate, the fall rate, the autocorrelation time, the base level stability, the low pulse duration, and the average date of the minimum head. To evaluate how well a modeled signature compares to the measured signature, an evaluation criterion (E_n) for each individual signature is calculated in the fourth step. The following evaluation criterion (adapted from Euser et al. 2013) is used here:

$$E_n = \max\left(0, 1 - \left|1 - \frac{S_n(h_{\text{sim}})}{S_n(h_{\text{obs}})}\right|\right) \tag{1}$$

where if E_n equals one, the simulated and measured signatures are equal and the model mimics perfectly the observed behavior. The lower bound of the evaluation criterion is set to zero, as simulations with values below zero either grossly over- or underestimate the signatures and indicate no skill of the model to reproduce the signature. Importantly, this helps the inter-comparison of performance between signatures.

The fifth and final step is then to aggregate the individual evaluation criteria into a single metric that is applied in the evaluation and selection process. A common approach is to use the Euclidean Distance (E_d, see, for example, Hrachowitz et al. 2014), computed as follows:

$$E_d = \sqrt{\frac{\sum_{n=1}^{N}(1 - E_n)^2}{N}} \tag{2}$$

where N is the number of signatures used in the evaluation. The performance E_d ranges between zero and one, where a value of one indicates a perfect model, that is, all modeled signatures mimic the measured signatures perfectly. Both the individual evaluation criteria (Equation 1) and the aggregate scores (Equation 2) can be used for model evaluation.

Table 2 shows the results of this analysis for the data shown in Figure 5. The value of a common goodness-of-fit metric, the Nash-Sutcliffe efficiency (NSE), is provided as well for comparison. The result shows that none of the models can perfectly mimic all the selected signatures simultaneously. Most signatures are, however, satisfactorily modeled by one or more of the models. Interestingly, the model with the lowest performance in terms of NSE (M2), scores best on the aggregate metric for the signatures (E_D), and many of the individual signatures.

The information in Table 2 can be used to learn about the causes of model defects, by relating the signatures to hydrological processes (i.e., the fall rate can be related to processes causing declines), or to select the more adequate model for the modeling purpose (i.e., if we are interested in the duration of a drought event). For example, the rise rate is best approximated by M2, while the fall rate is best approximated by M1. The fall rate is impacted by groundwater evaporation, which may cause a faster decline of the head. This process is missing from model M2, causing a lower fall rate. Interestingly, the average day of the year of the minimum head is perfectly mimicked by the M2 model, while the absolute value is not (see inset plot in Figure 5). Model M1 scores low on the base flow stability, caused by too many head variations due to an instant head response to precipitation. Models M2 and M3 perform better in this regard. This example shows how signatures can be used to identify potential routes for enhancing the models.

Potential Applications and Challenges of Signatures

Groundwater signatures can contribute to solving many common tasks of groundwater hydrologists. Below, we discuss several (potential) applications of groundwater signatures for the analysis and modeling of groundwater data. The different use cases are subdivided into three categories: (1) characterization and clustering, (2) improved understanding, and (3) groundwater modeling. The discussion presented below serves as a source of inspiration for incorporating groundwater signatures in subsequent research.

Characterization and Clustering

The first category of applications is to apply signatures for the purposes of characterization, comparison, and clustering (or classification) of groundwater systems. In this application, a set of signatures is computed for each head time series, which may then be compared and clustered. The example in the previous section illustrated how this is done for the national monitoring network in Switzerland. This was also done, for example, in Heudorfer et al. (2019) and Wunsch et al. (2022) for monitoring networks in Germany. This method may be preferred over methods directly using the head time series (see, e.g., Haaf and Barthel 2018), as they do not necessarily need overlapping time series and focus on groundwater behavior rather than absolute values. Giese et al. (2020) showed how signatures can distinguish between local, intermediate, and regional groundwater flow systems. Challenges here lie in the selection of the signatures used for characterization and the clustering algorithm used for the clustering. Future studies should systematically compare the outcomes of different approaches.

Improved Understanding

The second category of applications of groundwater signatures is their use to increase our understanding of the observed groundwater dynamics and the underlying processes causing these different dynamics (see also McMillan 2020). For this purpose, the signature values may be related to other observable properties, such as the climatic (e.g., temperature and precipitation patterns) and physiographic descriptors (e.g., depth to water table, altitude), that may explain part of the dynamics. Haaf et al. (2020) already showed that (perhaps unsurprisingly) clear relationships between many of such variables exist. These relationships can, for example, help us better understand how different groundwater systems respond to climatological changes.

Signature values can change over time, both due to system changes and changes in stresses on the groundwater. By computing signatures over different periods and comparing the values, signatures may be used to detect hydrological change. This type of application was illustrated in Barthel et al. (2021) for changing hydrological behavior due to dam developments, an example of human intervention in the hydrological system. Such analyses may also reveal more subtle changes, for example, changes in the first peak due to changing snowmelt patterns (e.g., as done by Bard et al. 2015, for streamflow data). This type of application may help uncover the impact of climate change on changing groundwater dynamics.

Groundwater Modeling

The third and final category of application of signatures is their use in different stages of groundwater modeling. In the initial stage, signatures may help to determine appropriate model structure(s) or candidates. David et al. (2022) tested the use of signatures to identify the required model complexity (i.e., structure) for rainfall-runoff modeling. Similar exercises can be performed for groundwater, where signatures are used to identify important processes and related model structures. Studies testing multiple model structures, that is, comparing different recharge process conceptualizations in Pastas (e.g., Brakenhoff et al. 2022; Collenteur et al. 2023; Jemel.janova et al. 2023) could potentially benefit from such an approach. For this purpose, clear relationships between the physiographic and climatic descriptors and the signatures need to be established. Additionally, signatures could enhance initial parameter estimates such as the hydraulic conductivity or storativity in numerical groundwater models or the gain in lumped parameter models. For this purpose, it needs to be explored if consistent empirical relationships exist between parameters and signatures. Promising results in this direction were found in Collenteur (2022, Chapter 5) for lumped parameter models. Such relationships can then be exploited to predict initial parameter values from signature values.

After a groundwater model is developed, it can be calibrated to groundwater signatures. An example of using signatures for calibration of streamflow models can be found in Kavetski et al. (2018). After model calibration, signatures may be used to evaluate model performance and to select a model if alternative models are available. This is already a relatively common approach in rainfall-runoff modeling (see, e.g., Euser et al. 2013; Viglione et al. 2013; Schaefli 2016). In both of these applications, the modeler may select the signatures that represent groundwater behavior that is important for the research question. One could thus adapt the calibration, evaluation, or selection criteria to obtain the best model for a specific purpose.

A final application of signatures in modeling groundwater dynamics is the prediction of groundwater dynamics in unmonitored aquifers, as proposed by Barthel et al. (2021). An initial proof-of-concept study from Haaf et al. (2023) showed potential for this method to predict heads in unmonitored (parts of) aquifers. In this approach, signatures (of the head duration curve) are used to characterize groundwater systems and predict groundwater dynamics based on similarity between aquifers. We refer to Haaf (2020) and Barthel et al. (2021) for detailed descriptions of the proposed approach.

Challenges and Limitations

One of the major challenges of using groundwater signatures lies in the selection of the signatures applied in an analysis. A good understanding of the meaning of the individual signatures is required to select (a set of) signatures that are focused on a specific part of the hydrograph or behavior. Table 1 and the documentation of the Pastas signature module and the references therein can be used for this purpose and help to select useful signatures. For larger sets of signatures, the correlations between signatures should be considered to ensure that the selection is not skewed toward one part of the behavior. Here, it is worth noting that these relationships may be nonlinear, requiring more complex methods such as singular value decomposition to uncover them and select uncorrelated signatures (Heudorfer et al. 2019; Nolte et al. 2024). More research is required into how to select signatures and which signatures are best to quantify specific parts of the hydrograph behavior.

The current set of signatures is primarily focused on signals at a lower frequency and more natural systems. For investigations into head time series characterized by higher frequency signals (i.e., daily pumping variations, tides), the applicability of the current set of signatures is probably limited, although this has not (yet) been tested. In the presence of such signals, another option would be to use standard capabilities of Pastas to filter out the contribution of higher frequency signals and analyze the signatures of the head fluctuations of the natural system. To use groundwater signatures to quantify the behavior of time series with high-frequency signals, the more interesting option is, however, to develop new signatures that are applicable to such time series and signals.

Finally, most published applications of groundwater signatures were aimed at and have shown success for larger scales, for example, regional scale hydrogeologic characterization, jointly in a large data set. In this study, we have provided an example of use for groundwater model evaluation that is useful for local scale modeling. However, more studies are needed that demonstrate how the concept of signatures can be useful on the local scale, where the core work of many hydrogeologists lies.

Concluding Remarks

In this paper, the Pastas groundwater signatures module is introduced. The module currently contains 31 fully documented and tested signatures. This number is intended to grow over time. Several tests were conducted and are available to test the identifiability, robustness, discriminatory power, and consistency of these signatures. The results of this analysis (Figures 2 and 3) may be used to make informed decisions on signatures selections in future studies. The tests indicate that many of the signatures perform well when tested on head data with different characteristics, but some are sensitive to measurement frequencies, time series lengths, and historic periods.

We believe groundwater signatures have many potential applications in groundwater studies. To encourage others to explore the use of signatures, two examples were provided to (1) cluster head time series and (2) evaluate groundwater models using signatures. Many more (potential) use cases were identified and discussed, and the Pastas signatures module is aimed at supporting future research into their use in the groundwater field. The module is part of the open-source Pastas software and freely available under the MIT license. We welcome contributions and feedback to the module from the open-source community, as well as newly developed groundwater-specific signatures. We highlight the need to develop a good signature for quantifying recession behavior, a signature that is of particular interest to our community and beyond, but remains difficult to robustly estimate in an automated way.

Acknowledgments

The authors thank the Executive Editor Charles Andrews and Jonathan Kennel for their constructive comments and insights that helped to further improve the manuscript. We thank the Swiss Federal Office for the Environment for providing the hydraulic head data for Switzerland. The research contributes to grant TRV2019/45670 awarded by the Swedish Transport Administration (Trafikverket).

Conflict of Interest

The authors declare no potential conflict of interest.

Data Availability Statement

All code and data used to generate the numbers and figures in this manuscript is available from the Zenodo repository: https://doi.org/10.5281/zenodo.10604325. The Pastas signatures module is available from the Pastas software as open-source code under the MIT License (Collenteur et al. 2019b).

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Data S1. The supporting information shows more detailed information on two guidelines: **Identifiability** showing the impact of added errors and smoothing on various signatures; **Robustness** with figures illustrating the effects of different measurement frequencies, time windows, and time series lengths on signature values.

References

Bard, A., B. Renard, M. Lang, I. Giuntoli, J. Korck, G. Koboltschnig, M. Janža, M. d'Amico, and D. Volken. 2015. Trends in the hydrologic regime of Alpine rivers. *Journal of Hydrology* 529: 1823–1837. https://doi.org/10.1016/j.jhydrol.2015.07.052

Barthel, R., E. Haaf, M. Giese, M. Nygren, B. Heudorfer, and K. Stahl. 2021. Similarity-based approaches in hydrogeology: Proposal of a new concept for data-scarce groundwater resource characterization and prediction. *Hydrogeology Journal* 29, no. 5: 1693–1709. https://doi.org/10.1007/s10040-021-02358-4

Brakenhoff, D.A., M.A. Vonk, R.A. Collenteur, M. Van Baar, and M. Bakker. 2022. Application of time series analysis to estimate drawdown from multiple well fields. *Frontiers*

in Earth Science 10: 907609. https://doi.org/10.3389/feart.2022.907609
Collenteur, R.A. 2022. Improved time series analysis of groundwater data through open-source software and better process representations. Dissertation, Universität Graz. https://resolver.obvsg.at/urn:nbn:at:at-ubg:1-181228 (accessed August 19, 2024).
Collenteur, R.A., C. Moeck, M. Schirmer, and S. Birk. 2023. Analysis of nationwide groundwater monitoring networks using lumped-parameter models. *Journal of Hydrology* 626: 130120. https://doi.org/10.1016/j.jhydrol.2023.130120
Collenteur, R.A., M. Bakker, R. Caljé, S.A. Klop, and F. Schaars. 2019a. Pastas: Open source software for the analysis of groundwater time series. *Groundwater* 57, no. 6: 877–885. https://doi.org/10.1111/gwat.12925
Collenteur, R.A., M. Bakker, R. Caljé, and F. Schaars. 2019b. Pastas: open-source software for the analysis of hydrogeological time series. https://zenodo.org/record/3252035#.XQz7oy2B1TY (accessed August 19, 2024).
David, P.C., P.L.B. Chaffe, V.B.P. Chagas, M. Dal Molin, D.Y. Oliveira, A.H.F. Klein, and F. Fenicia. 2022. Correspondence between model structures and hydrological signatures: A large-sample case study using 508 Brazilian catchments. *Water Resources Research* 58, no. 3: e2021WR030619. https://doi.org/10.1029/2021WR030619
Euser, T., H.C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H.H.G. Savenije. 2013. A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences* 17, no. 5: 1893–1912. https://doi.org/10.5194/hess-17-1893-2013
Giese, M., E. Haaf, B. Heudorfer, and R. Barthel. 2020. Comparative hydrogeology – Reference analysis of groundwater dynamics from neighbouring observation wells. *Hydrological Sciences Journal* 65, no. 10: 1685–1706. https://doi.org/10.1080/02626667.2020.1762888
Haaf, E. 2020. Towards prediction in ungauged aquifers – Methods for comparative regional analysis. PhD thesis, University of Gothenburg.
Haaf, E., M. Giese, T. Reimann, and R. Barthel. 2023. Data-driven estimation of groundwater level time-series at unmonitored sites using comparative regional analysis. *Water Resources Research* 59, no. 7: e2022WR033470. https://doi.org/10.1029/2022WR033470
Haaf, E., M. Giese, B. Heudorfer, K. Stahl, and R. Barthel. 2020. Physiographic and climatic controls on regional groundwater dynamics. *Water Resources Research* 56, no. 10: e2019WR026545. https://doi.org/10.1029/2019WR026545
Haaf, E., and R. Barthel. 2018. An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs. *Journal of Hydrology* 559: 222–237. https://doi.org/10.1016/j.jhydrol.2018.02.035
Heudorfer, B., E. Haaf, K. Stahl, and R. Barthel. 2019. Index-based characterization and quantification of groundwater dynamics. *Water Resources Research* 55, no. 7: 5575–5592. https://doi.org/10.1029/2018WR024418
Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H.H.G. Savenije, and C. Gascuel-Odoux. 2014. Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research* 50, no. 9: 7445–7469. https://doi.org/10.1002/2014WR015484
Jemeljanova, M., R.A. Collenteur, A. Kmoch, J. Bikše, K. Popovs, and A. Kalvāns. 2023. Modeling hydraulic heads with impulse response functions in different environmental settings of the Baltic countries. *Journal of Hydrology: Regional Studies* 47: 101416. https://doi.org/10.1016/j.ejrh.2023.101416
Kavetski, D., F. Fenicia, P. Reichert, and C. Albert. 2018. Signature-domain calibration of hydrological models using approximate Bayesian computation: Theory and comparison to existing applications. *Water Resources Research* 54, no. 6: 4059–4083. https://doi.org/10.1002/2017WR020528
McKinney, W. 2010. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, ed. S.v.d. Walt, and J. Millman, 51–56. Austin, Texas: SciPy.
McMillan, H.K. 2021. A review of hydrologic signatures and their applications. *WIREs Water* 8, no. 1: e1499. https://doi.org/10.1002/wat2.1499
McMillan, H.K. 2020. Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes* 34, no. 6: 1393–1409. https://doi.org/10.1002/hyp.13632
McMillan, H., G. Coxon, R. Araki, S. Salwey, C. Kelleher, Y. Zheng, W. Knoben, S. Gnann, J. Seibert, and L. Bolotin. 2023. When good signatures go bad: Applying hydrologic signatures in large sample studies. *Hydrological Processes* 37, no. 9: e14987. https://doi.org/10.1002/hyp.14987
McMillan, H., I. Westerberg, and F. Branger. 2017. Five guidelines for selecting hydrological signatures. *Hydrological Processes* 31, no. 26: 4757–4761. https://doi.org/10.1002/hyp.11300
Nolte, A., E. Haaf, B. Heudorfer, S. Bender, and J. Hartmann. 2024. Disentangling coastal groundwater level dynamics on a global data set. *Hydrology and Earth System Sciences Discussions* 28: 1215–1249. https://doi.org/10.5194/hess-28-1215-2024
Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12: 2825–2830.
Rinderer, M., H.J. van Meerveld, and B.L. McGlynn. 2019. From points to patterns: Using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics. *Water Resources Research* 55, no. 7: 5784–5806. https://doi.org/10.1029/2018WR023886
Schaefli, B. 2016. Snow hydrology signatures for model identification within a limits-of-acceptability approach. *Hydrological Processes* 30, no. 22: 4019–4035. https://doi.org/10.1002/hyp.10972
Viglione, A., J. Parajka, M. Rogger, J.L. Salinas, G. Laaha, M. Sivapalan, and G. Blöschl. 2013. Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences* 17, no. 6: 2263–2279. https://doi.org/10.5194/hess-17-2263-2013
Vonk, M.A., R.A. Collenteur, S. Panday, F. Schaars, and M. Bakker. 2024. Time series analysis of nonlinear head dynamics using synthetic data generated with a variably saturated model. *Groundwater* 62, no. 5: 748–760. https://doi-org/10.1111/gwat.13403
Wunsch, A., T. Liesch, and S. Broda. 2022. Feature-based groundwater hydrograph clustering using unsupervised self-organizing map-ensembles. *Water Resources Management* 36, no. 1: 39–54. https://doi.org/10.1007/s11269-021-03006-y