https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

Disentangling coastal groundwater level dynamics on a global data set

Annika Nolte[1], Ezra Haaf[2], Benedikt Heudorfer[3], Steffen Bender[1], Jens Hartmann[4]

[1]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Hamburg, Germany
[2]Chalmers University of Technology, Gothenburg, Sweden
[3]Karlsruhe Institute of Technology (KIT), Institute of Applied Geosciences, Hydrogeology, Karlsruhe, Germany
[4]Institute for Geology, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany

Correspondence to: Annika Nolte (annika.nolte@hereon.de)

Abstract

This study aims to identify common hydrogeological patterns and to gain a deeper understanding of the underlying similarities and their link to physiographic, climatic, and anthropogenic controls of coastal groundwater. The most striking aspects of GWL dynamics and their controls were identified through a combination of statistical metrics, calculated from about 8,000 groundwater hydrographs, and pattern recognition, classification, and explanation using machine learning techniques and SHapley Additive exPlanations (SHAP). Overall, four different GWL dynamics patterns emerge, independent of the different seasons, time series lengths, and periods. We show in this study that similar GWL dynamics can be observed around the world with different combinations of site characteristics, but also that the main factors differentiating these patterns can be identified. Three of the identified patterns exhibit high short-term and interannual variability and are most common in regions with low terrain elevation and shallow groundwater depth. Climate and soil characteristics are most important in differentiating these patterns. This study provides new insights into the hydrogeological behavior of groundwater in coastal regions and guides systematic and holistic groundwater monitoring and modelling, motivating to consider various aspects of GWL dynamics when, for example, estimating climate-driven GWL changes – especially when information on potential controls is limited.

1. Introduction

When GWL dynamics are tracked over time using groundwater hydrographs, the quantitative status of groundwater resources can be determined. Groundwater resources are tapped and measured locally, and GWL dynamics respond to processes in their immediate local and regional environment, such as groundwater recharge rates, groundwater flow and pumping, and seawater intrusion (SWI) in coastal aquifers. Hence, they are strongly location-bound, and there is "a need for groundwater assessments at the field level" (United Nations, 2022). To this end, it is common practice in local studies to incorporate direct information from the groundwater system combined with expert knowledge of potential controls into numerical, statistical, or machine learning models (e.g. in Knowling et al., 2015, Güler et al., 2012 and Lee et al., 2019). However, comprehensive knowledge of aquifer processes at the local scale is often lacking, posing even greater challenges for regional, continental, or global assessments of groundwater systems. Such larger scales are important, for example, concerning large or transboundary aquifer

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

systems, global virtual water trade, and international frameworks such as the UN Sustainable Development Goals (Donnelly et al., 2018; Nimmo et al., 2021). A systemic understanding of aquifer properties is key to sustainable groundwater management and governance (Guppy et al., 2018; Elshall et al., 2020). Generalized scientific understanding of processes currently relies heavily on aggregating and upscaling local knowledge of groundwater's bidirectional connections with the land surface and submarine processes (Smerdon, 2017; Lall et al., 2020; Gleeson et al., 2020), however, is often lacking direct observation-based evidence.

While it is largely unclear how GWL dynamics compare at large scales, the classification of GWL time series applied on such scales holds the potential to provide valuable insights into hydrogeological similarity (Barthel et al., 2021). For instance, these insights can prove valuable in assessing the coherence of large-scale process-based models by focusing on similarities in patterns and spatial trends, rather than solely relying on the magnitude of aggregated errors, effectively mitigating the commensurability problem (Gleeson et al., 2021). The present study follows this idea and is motivated by (a) a large amount of GWL data available today – although not yet centralized and freely accessible for many regions of the world (United Nations, 2022) – and (b) the advancement of data-driven methods in earth system sciences, which are not only capable of finding patterns unseen by humans but are also increasingly capable of explaining them (Reichstein et al., 2019; Shen et al., 2018). Previous studies have successfully applied inductive classification approaches (Olden et al., 2012), synthesizing different aspects of GWL dynamics at local to regional scales to investigate physiographic, climatic, and anthropogenic controls of GWL dynamics (Giese et al., 2020; Haaf et al., 2020; Ascott et al., 2020; Wunsch et al., 2021; Sorensen et al., 2021), the function of groundwater in ecosystems (Martens et al., 2013) or runoff processes (Rinderer et al., 2019). Other studies have pointed out that understanding processes is limited when only the long-term mean or trend of GWL dynamics is used (Lischeid et al., 2021; Gleeson et al., 2021; Baulon et al., 2022a). In this study, we aim at answering the following research questions:

1. How is hydrogeological similarity observed on a global scale, and what are the implications of scaling on similarity patterns?
2. What are the key controlling factors for GWL dynamics patterns on a global scale, and how do they explain the variations observed at smaller scales?
3. What are the current opportunities and barriers to deriving generalizations of dynamics-control relationships in groundwater using data-driven approaches?

This study accesses a newly compiled large and diverse dataset of about 8,000 GWL time series of coastal aquifers from five continents. Information from the compiled and pre-processed time series (Sect. 2.1) was analyzed holistically using a set of previously defined statistical metrics from Heudorfer et al. 2019 computed from individual groundwater hydrographs (Sect. 2.2), thereby reducing their dimensionality (Wang et al., 2006). Sect. 3.1 derives GWL dynamics patterns in coastal regions by using unsupervised clustering techniques to group GWL time series based on information derived from these metrics (Sect. 2.3). The importance and vulnerability of coastal groundwater, which serves as a vital freshwater resource for ecosystems and large coastal communities with increasing water demands from groundwater (Oude Essink et al., 2010; Famiglietti, 2014;

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Ferguson and Gleeson, 2012; Moosdorf and Oehler, 2017), prompts this study to focus on disentangling GWL dynamics in coastal regions. To explain these findings, independent data encompassing various potential controls of GWL dynamics and associated surface processes were obtained from global map products (Sect. 2.4) and were used to predict clusters in a Random Forest (RF) classification task (Sect. 2.5) with the results being presented in Sect. 3.2. SHAP values were calculated for the RF model (Sect. 2.6). These are used to explain relationships between predicted clusters and descriptive features in Sect. 3.3. We then discuss hydrogeological similarity results and the scaling effect (Sect. 4.1), evaluate derived explanations for GWL dynamics patterns (Sect. 4.2), and contextualize our findings within the framework of a case study region (Sect. 4.3).

2. Data and methods

The design of this study consists of the calculation of statistical metrics (hereafter referred to as *indices*) using compiled GWL time series, and the clustering of GWL dynamics by applying unsupervised algorithms on the calculated indices. Subsequently, the clustering result that best differentiates characteristic groups of GWL dynamics was fed into an RF classifier together with local and regional natural and anthropogenic characteristics (hereafter referred to as *attributes*) from global map products that are potentially related to GWL dynamics. SHAP values were derived to understand the salient controls of each group of GWL dynamics. Unless stated otherwise, data pre-processing, modelling, and visualizations were done using Python (Python Software Foundation, 2021).

2.1 GWL time series and pre-processing

The GWL time series data set analyzed in this study was compiled from national or state governmental agencies listed in Table A1, in the years 2019 to 2022. Although attempts were made to find data for all coastal regions in the world, data collection focused on regions with a long coastline and large quantities of digitally available data. The wells from the compiled data set are mainly located in northwestern Europe (Belgium, Denmark, France, Germany, Ireland, Netherlands, and Sweden), but also in Australia, South Africa, Brazil, Canada, and the United States. The wells are part of the GWL monitoring networks of the respective countries, but anthropogenic impact on GWL dynamics cannot be ruled out for all wells. In addition, wells may be affected by SWI and thus have variable density (Post et al., 2007). However, density gradients are not analyzed in this study, because our focus is on absolute variations of the GWL. The GWL time series were selected based on the following criteria:

a) Availability of secondary information (in convertible units for homogenization):
   - Latitude and longitude of the well location (with variable accuracy)
   - Reference vertical point/datum to measure the GWL depth
   - Date of observation

b) Well location and situation:

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

- Selection of wells located at the global coasts: Well is located within 100 km distance of the shoreline. This follows the definition of Martínez et al. (2007) and Mangor et al. (2017).
- Selection of shallow aquifers with mean GWLs less than 100 m below ground surface. It is important to note that this criterion is used as a directional guideline rather than an exact marker, as there is uncertainty of a few meters in the absolute groundwater depth of observations referenced to sea level that were converted to ground surface reference with elevation data from the Shuttle Radar Topography Mission (SRTM) (Kulp and Strauss, 2018; Rodriguez et al., 2006; source-specific information in the Supplement).

c) Temporal availability:

- Data must be available for at least four complete calendar years during the period: 1979-2019.
- Time series must not have more than ten percent missing observations with data gap lengths of no larger than two weeks after aggregation to weekly time steps.

Further source-specific pre-processing steps were necessary because of the very diverse data presentation, for example, regarding format, labels, and units. The reader is referred to the Supplement for notes on groundwater data collection and pre-processing. Potential anomalies or change points of the individual time series that indicate human activities, measurement, or reporting errors were removed using Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Ester et al., 1996) combined with - due to the large number of time series - visual inspection of suspicious time series (Fig. A1). In line with recent publications, a systematic and additional visual inspection of groundwater time series was found to be beneficial (Barthel et al., 2022; Berendrecht et al., 2022; Lehr and Lischeid, 2020; Retike et al., 2022). Subsequently, temporal availability criteria were checked once again, with remaining data gaps being linearly interpolated and time series were transformed to the 0-1 scale for calculating indices.

2.2 Indices

We calculated a total of 45 indices for all GWL time series using an unpublished R package (R Core Team, 2021) that was developed with the study conducted by Heudorfer et al. (2019). These indices statistically aggregate and describe various aspects of groundwater hydrographs, its *structure* (e.g., interannual fluctuation heights and changes), *distribution* (e.g., alignment of the GWL with upper or lower fluctuation limits), and its *shape* (e.g., steepness of hydrograph increases and decreases). For more detailed descriptions of the used indices (Table B1), we refer to Heudorfer et al. (2019). In addition, there are other promising signature-based time series characterizations for hydroclimatic time series, as selected and explored in Papacharalampous and Tyralis (2022), Papacharalampous et al. (2023) and Wunsch et al. (2021).

For our approach, the indices were computed from weekly aggregated time series with a length of at least four years (Sect. 2.1), which is at the lower end of the period of four to eight years recommended in Heudorfer et al. 2019. We decided to use short periods to enable a larger pool of time series to be taken into account for the global similarity analysis. The calculated set of indices transforms groundwater information into 45 static metrics. The resulting index value ranges were compared to

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

those of Heudorfer et al. (2019) for quality control (Fig. B1). As an outlier treatment, time series with index values in the outermost 0.001 % of the index value distribution (4 * σ) were discarded.

2.3 Cluster analysis

Clustering of the indices was performed to find a generalized but robust representation of differing GWL dynamics in coastal regions. No specific indices were selected a priori to capture GWL dynamics holistically. This decision was made, because indices differ in their ability to describe GWL dynamics depending on the flow system and the groundwater regime (Heudorfer et al., 2019; Giese et al., 2020; Haaf et al., 2020). For a similar reason, Papacharalampous et al. (2023) recommended using a large variety of time series characteristics for similarity analysis in hydrology.

A Principal Component Analysis (PCA) was applied to index values after standardization to avoid redundancies among the 45 indices (Heudorfer et al., 2019; Fig. B2) and to reduce the number of features for clustering. The number of Principle Components (PCs) to be retained was set based on the scree plot and variance explained, yielding a dimensionality-reduced dataset, where indices are mapped as score values on the significant PCs. This was followed by a systematic search for the best aggregation on the scores using three different unsupervised clustering algorithms – agglomerative hierarchical clustering using the Ward criterion, k-means, and Gaussian mixture. Unsupervised methods aim to learn more about the internal dependencies among the explanatory variables (Bergen et al., 2019), meaning that no expectations regarding the number or patterns of clusters were pre-set. Instead, evaluation metrics (Rousseeuw, 1987; Caliński and Harabasz, 1974; Davies and Bouldin, 1979) were used to find the best arrangement of data points into clusters via optimizing within-cluster and between-cluster similarity and dissimilarity.

2.4 Attribute data

Understanding GWL dynamics requires integrated information from its manifold controls. These are typically represented by attribute data from climate, landuse and landcover, soil and lithology, surface waters, topography, subsurface structures, and (hydro)geology and anthropogenic activity (Moeck et al., 2020; Rajaee et al., 2019; Díaz-Alcaide and Martínez-Santos, 2019). For river catchments, such datasets have already been collected for the regional scale (Addor et al., 2017; Klingler et al., 2021) and the global scale, most notably the HydroATLAS (Linke et al., 2019). However, to date, there are no global datasets available that encompass the above-mentioned attributes specifically for groundwater studies (Haaf et al., 2020). Therefore, attributes mainly describing natural and anthropogenic characteristics and processes at the surface were collected from a variety of recent highest-resolution datasets available at the global scale (Table 1). These were used as proxies for global groundwater-specific datasets in this study.

While river catchments are typically well-defined by topography, subsurface catchments are often complex and vary on small scales (Vahdat-Aboueshagh et al., 2021). Despite recent efforts to develop a similar approach for calculating watersheds with

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

a better representation for groundwater (Nölscher et al., 2022; Huggins et al., 2023), at the moment there is no best-accepted method for groundwater-relevant delineation of various surface attributes on large scales. One approach to defining the contributing area of a groundwater well is to rely on the immediate vicinity and place a circular buffer of selected diameter around the monitoring site (Johnson and Belitz, 2009; Knoll et al., 2019). Using this approach, attributes were extracted as averages from a 500 m buffer. The prepared attributes are of numerical and categorical type and contain information from multiple periods that have been unified where possible. The spatial resolution of the underlying source data is less than 30 arcseconds, which is about 1000 meters, except for hydrological data, which is derived from grids with a spatial resolution of only 1800 arcseconds (55 km).

Table 1 List (alphabetically by attribute category) of numeric (num) and categorical (cat) attributes used to relate GWL dynamics to controls and processes in this study. Single or multidimensional raster data and spatial vector data were downloaded and processed for well locations as follows (*): [1] Average (median or majority) raster values were extracted for buffer geometries using the zonal statistics tool in ArcGIS Pro (Esri, 2022) using a cell size of 250 m. [2] Earth Engine Python API was accessed to directly calculate median raster values for buffer geometries and data download [3] Average (median or majority) values from geospatial vector data were extracted for buffer geometries using the spatial join tool in ArcGIS Pro. [4] Distance was calculated between well points and polylines using the near tool in ArcGIS Pro. [5] Multidimensional raster were aggregated to a single raster layer that presents the median or majority of superimposed raster cells of the period from 1981 to the latest available temporal dimension (make multidimensional raster tool and raster aggregation tool in ArcGIS Pro). [6] Calculated the median of multiple raster files using the cell statistics tool in ArcGIS Pro. [6] Rivers that have a catchment area of at least 10 km² or an average river flow of at least 0.1 m³/sec, or both.

category	attribute name	short description	unit (converted)	data type	period (averaged)	source temporal resolution	source spatial resolution	reference	*
anthropogenic	population_density	population density	persons km^{-2}	num	2000	annual	30 arc sec	CIESIN (2018)	1
	irrigation	irrigated area serviced by groundwater	%	num	2000	multi-annual	5 acr min	Siebert et al. (2015)	1
	groundwater_abstraction	actual net abstraction from groundwater	kg m^{-2} a^{-1}	num	1981-2016	monthly	1800 arc sec	Müller Schmied et al. (2020)	5,1
climate	temperature_mean	mean annual daily mean air temperatures	°C	num	1981-2010	period average	30 arc sec	Karger et al. (2017)	1
	temperature_max	highest temperature of any monthly daily mean maximum temperature	°C	num	1981-2010	period average	30 arc sec	Karger et al. (2017)	1
	temperature_min	lowest temperature of any monthly daily mean maximum temperature	°C	num	1981-2010	period average	30 arc sec	Karger et al. (2017)	1
	precipitation	accumulated precipitation amount	kg m^{-2} a^{-1}	num	1981-2010	period average	30 arc sec	Karger et al. (2017)	1
	PET	evapotranspiration of reference crop (ET0)	mm day^{-1}	num	1970-2000	period average	30 arc sec	Trabucco and Zomer (2019)	1
	aridity	aridity index (high for more humid conditions, and low for more arid conditions)	-	num	1970-2000	period average	30 arc sec	Trabucco and Zomer (2019)	1

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

landuse and landcover	landuse	major landuse class	-	cat	1981-2019	annual	30 arc sec	Winkler et al. (2020)	5,1
	ecosystem	major terrestrial ecosystem class	-	cat	-	-	7.5 arc sec	Sayre et al. (2020)	1
soil and lithology	soc	soil organic carbon 0-5 cm mean	dg kg^{-1}	num	-	-	7.5 arc sec	Poggio et al. (2020)	2,3
	bdod	bulk density 0-5 cm mean	dg cm^{-3}	num	-	-	7.5 arc sec	Poggio et al. (2020)	2,3
	clay	clay content 0-5 cm mean	g kg^{-1}	num	-	-	7.5 arc sec	Poggio et al. (2020)	2,3
	sand	sand content 0-5 cm mean	g kg^{-1}	num	-	-	7.5 arc sec	Poggio et al. (2020)	2,3
	silt	silt content 0-5 cm mean	g kg^{-1}	num	-	-	7.5 arc sec	Poggio et al. (2020)	2,3
	lithology	major lithological class	-	cat	-	-	polygons	Hartmann and Moosdorf (2012)	3
	unconsol_sediment	unconsolidated sediments major class	-	cat	-	-	polygons	Börker et al. (2018)	3
	permeability	surface permeability	log(k) m^2	num	-		polygons	Huscroft et al. (2018)	3
surface waters	streamflow	mean annual river discharge	m^3 s^{-1}	num	1981-2015	annual	30 arc sec	Barbarossa et al. (2018)	5,1
	distance_coast	distance of well location to the coastline	km	num	-		polyline	Lehner and Grill (2013)	4
	distance_stream	distance of well location to the closest river	km	num	-		polyline	Lehner and Grill (2013)	4,6
	runoff	runoff from land (sum of runoff components)	kg m^{-2} a^{-1}	num	1981-2016	monthly	1800 arc sec	Müller Schmied et al. (2020)	5,1
	TWS	total water storage (sum of all water storage compartments)	kg m^{-2}	num	1981-2016	monthly	1800 arc sec	Müller Schmied et al. (2020)	5,1
topo-graphy and sub-surface depths	bedrock_depth	absolute depth to bedrock	cm	num	-	-	7.5 arc sec	Shangguan et al. (2017)	1
	wtd	depth to the shallow groundwater table	m	num	-	-	30 arc sec	Fan et al. (2013)	5,1
	elevation	elevation	m	num	1994	-	1 arc sec	Farr et al. (2007)	2,3
	slope	steepness of elevation surface	-	num	1994	-	1 arc sec	Farr et al. (2007)	2,3

175

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

2.5 Random Forest model

From n identified clusters (Sect. 2.3), the classification task was to assign the correct cluster from a set of attributes (Sect. 2.4). Well locations were considered for RF classification (Breiman, 2001) only when data for all attributes were available. Furthermore, to avoid multiple clusters for a single combination of attributes, the cluster with the highest frequency was selected. The instance was retained if the identified cluster constituted more than 50 % of all clusters associated with that specific combination of attributes. The categorical attributes landuse, ecosystem, unconsolidated sediment, and lithology were pre-processed for their use in the RF model using one-hot encoding.

The RF model was optimized and evaluated with a stratified split of the dataset into a group for training and optimization (80 % of the data) and a group for testing (20 % of the data), taking into account an imbalanced dataset, where not all clusters have equal numbers of wells assigned and where regions are not represented equally by the dataset. For model optimization, hyperparameters were tuned within a 5-fold cross-validation framework. RF is a tree-based method that contains structures for feature selection (Breiman, 2001). No explicit feature selection was applied during pre-processing or model optimization, driven by the objective of exploring the relative importance and effects of all potential controls as well as their collective impacts on GWL dynamics. Evaluation during model optimization and with the testing data of the optimized model was carried out using the model accuracy, calculated as the total number of correctly assigned classes divided by the sample number. In addition, the optimized model, i.e. the success of the classifier, was evaluated with metrics for the individual clusters: precision; recall; F1-score. Here, precision is the number of correctly-identified wells per cluster divided by all the times the model predicts that cluster. Recall is the number of correctly-identified wells per cluster divided by the total number of wells in that cluster. The F1-score is twice the multiplication of precision and recall (2 * precision * recall) divided by the sum of precision and recall. RF in general works reasonably well when using the default values of hyperparameters in common algorithms (Probst et al., 2019). However, performance improvements can still be achieved by tuning especially the number of trees that should be set high enough for robust results and the number of features considered for each time the tree is split. Up to 2,000 trees were tested in combination with either all features or their square root being considered when looking for the best split using two different splitting criteria (“Gini” impurity and “Entropy”: Breiman, 2001).

2.6 SHAP analysis

SHapley Additive exPlanations (SHAP) values (Lundberg et al., 2020) were used to investigate the feature attributions (feature importance and feature effects) to explain the RF classification. Such approaches of explainable machine learning have been rarely used, but are increasingly and successfully applied in hydrology (Worland et al., 2019; Yang and Chui, 2021; Wunsch et al., 2022; Liu et al., 2022; Haaf et al., 2023). The three-dimensional space of SHAP values for the multiclass problem of this study is given by the number of clusters, the number of samples, and the number of features, i.e. attributes. Larger SHAP values for a specific cluster correspond to a higher probability of the cluster label. SHAP values of the samples that are part of the test dataset were analyzed collectively to explain how attributes impact the classification of the model. This corresponds

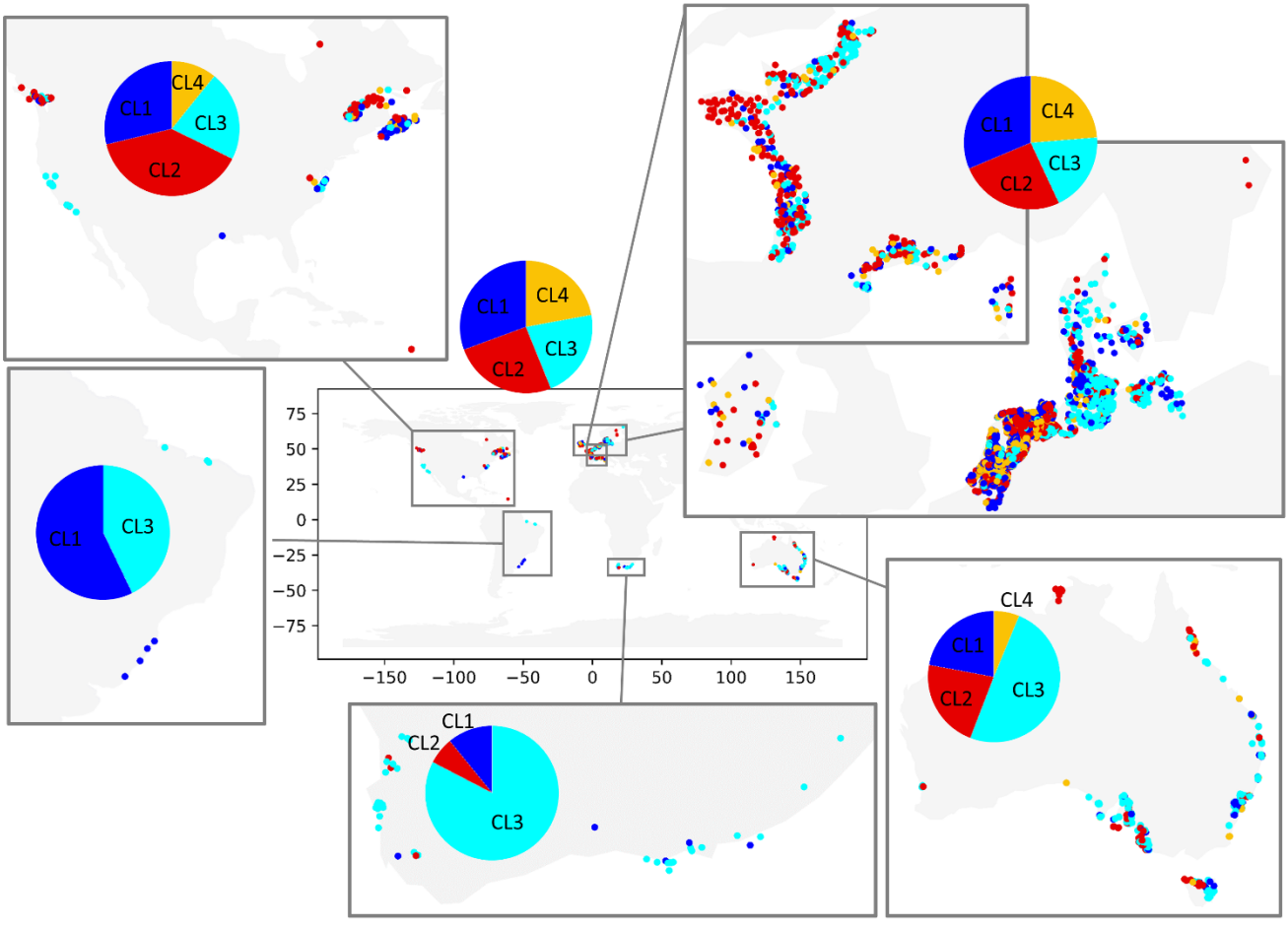
https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

to a global explanation of the model. For some analyses, SHAP values were combined for the attribute categories listed in Table 1. The SHAP values of individual instances, i.e. for well sites, were combined with regional information to discuss the extent to which the generalized GWL dynamics derived through the indices approach of this study can be explained in a case study region. When explaining controls of GWL dynamics from SHAP values, it is important to consider the interrelationships between the features. From the point of redundant features, the model is mainly trained on the feature with the most information in combination with other selected features. Hence variables that combine information included in other features are generally ranked higher and reduce the ranking of the other redundant features.

3. Results

3.1 Identification of groundwater level dynamics patterns



https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

Figure 1 Assignment of clusters to well locations globally (center) and in coastal regions in North America, Europe (enlarged map of France and northwestern Europe), Australia, South Africa, and Brazil. Due to the large scale, overlapping dots are not jittered, but well points are shuffled to prevent the same clusters from being constantly drawn over others.

Application of the time series selection criteria from Sect. 2.1 resulted in a set of 8,574 GWL time series for which 45 indices were calculated, respectively. After removing time series with extreme index values (Sect. 2.2), a dataset with indices for 7,888 time series remained. The value ranges for the different indices are comparable to those found by Heudorfer et al. (2019) (Fig. B1). Several indices showed a strong linear correlation (Fig. B2). From the PCA, five PCs represented 70% of the variability of the linearly combined indices, while the first three PCs already described more than half of the variability (Fig. C1). The top five PCs served as input for the three clustering algorithms. The three algorithms yielded similar results, where the highest cluster separation was achieved with three or four clusters (CL). K-means yielded the best cluster separation compared to hierarchical clustering and Gaussian mixture on all three evaluation criteria for cluster separation (Fig. C2). Comparing the cluster membership of wells, cluster composition based on k-means clustering is found to be quite similar to hierarchical clustering for most well sites (Fig. C3). The various characteristics of GWL dynamics can be assumed to exhibit complex and non-linear interlinkages in the multidimensional space.

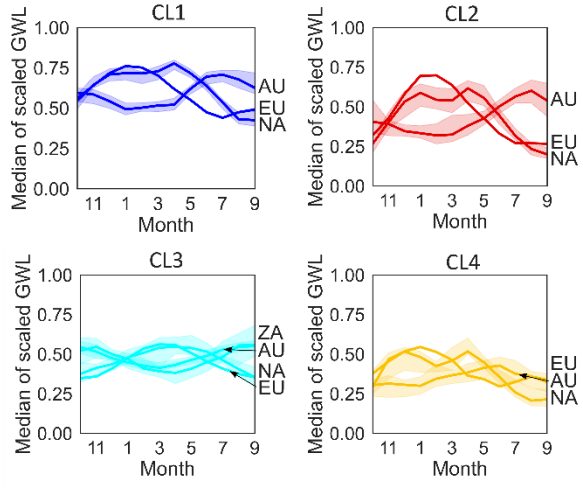


Figure 2 Annual hydrographs (median and 95% confidence interval) from averaged groundwater depths of scaled GWL time series for each month per cluster, calculated and plotted for different coastal regions in North America (NA), Europe (EU), South Africa (SA), and Australia (AU) when at least ten time series associated with the respective cluster were available.

Given the nature of the data set, the success of the different clustering approaches for disentangling GWL dynamics can be mainly attributed to the following two aspects. Firstly, it is plausible that the data distribution may not strictly follow Gaussian distributions because already individual indices show nonlinear behavior (Fig. B1; Haaf et al., 2020), limiting the effectiveness of Gaussian mixture models. In such cases, the characteristics of the dataset can favor clustering algorithms such as k-means and hierarchical clustering that are more flexible with regard to the range of cluster shapes and sizes. Secondly, multidimensional datasets with complex and non-linear interlinkages may often lack clearly defined and compact clusters (Campello et al., 2020), which becomes evident upon analyzing the scatter plot of the first three PCs (Fig. C3). This makes

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

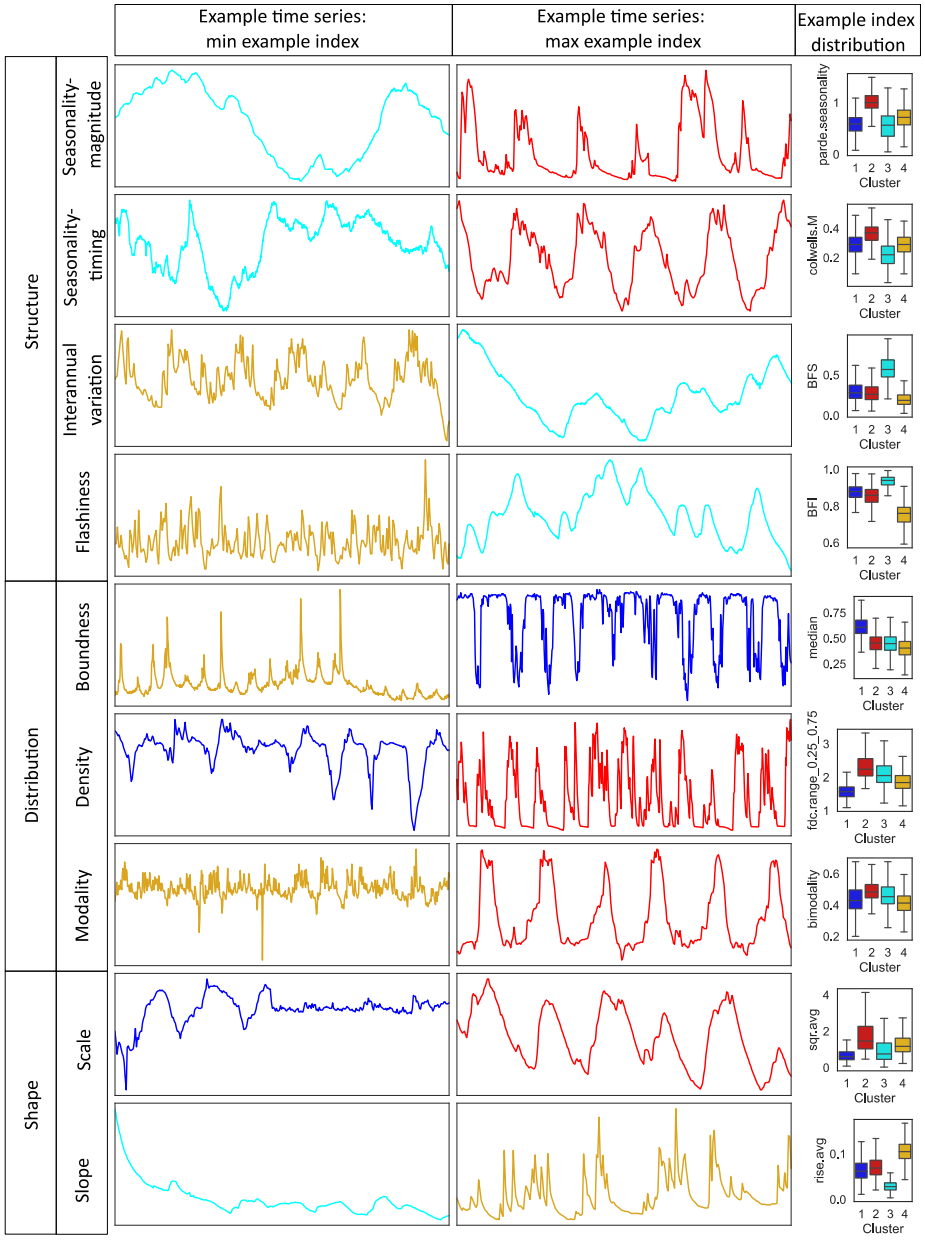


Figure 3 Example groundwater time series from the European continent illustrate typical GWL time series where one aspect from each component of the Heudorfer et al. (2019) typology of GWL dynamics (Table B1) is weak and strong, using an example index from that component that is displayed showing within-cluster variability (see Fig. C4 for within-cluster variability for all indices).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

The predominant cluster differs between the coastal regions represented in this study. In regions with a high density of wells, usually all clusters are present (Fig. 1). Therefore, similar patterns of GWL dynamics are found in coastal aquifers across the globe. Regional spatial patterns, i.e. geographically concentrated clusters, are visible for South Africa and Australia where most wells belong to cluster CL3 and within Europe, where most of the wells in the dataset are located. For example, in Germany wells assigned to CL3 are mostly located around the Baltic Sea and further inland, while the other clusters are pronounced along the exposed west coast. In the Netherlands and Belgium, most wells belong to CL1 and CL2. Most wells in north-eastern France are assigned to CL3, whereas most wells in north-western France are assigned to CL2. Each cluster represents a distinct pattern of GWL dynamics, regardless of shifted or opposing seasons in different regions, as shown in Fig. 2, where, for example, the double hump in the average annual hydrograph of North America with time series mainly in the upper latitudes could be attributed to snowmelt and is found in all clusters. Patterns are the most precise (smallest confidence intervals) for the European dataset.

GWLs in three out of four clusters can be located either mostly at their upper (CL1) or lower (CL4) boundary during the year with high GWLs in winter and spring to low GWLs in summer, or the GWL fluctuates around its annual median during the year with high interannual amplitude (CL2). In contrast, are the GWLs of CL3, where the maximum of the water levels is reached later in the year.

The clusters can be distinguished with respect to multiple features of the hydrograph, expressed by indices. Thus, the salient pattern of each cluster is defined by several indices, not by an individual one. However, some individual indices allow to separate clusters or groups of clusters better than the others. Fig. 3 shows how different aspects of GWL dynamics, each described by multiple indices (Table B1), contribute to disentangling patterns of GWL dynamics. Many indices of multiple components clearly separate clusters concerning the hydrograph *structure*, including the regularity of the seasonal amplitude (*parde.seasonality*) that is small for CL1 and CL3 and largest for CL2, the size of the global amplitude of the unscaled GWL that is comparatively small for CL4, the manifestation of interannual variation (e.g. from *BFS* that implies that interannual variation is rather stable in CL3 and unstable in CL4) and different order of flashiness (e.g. from *BFI* where the baseflow component is large for CL3 and small for CL4). From the time series *distribution* aspect, GWL dynamics from CL1 can be characterized as upper bounded, while GWL dynamics from CL4 are slightly lower bounded (e.g. from *median*). Dispersion over the range of GWLs (e.g. from the range of duration curve (*fdc.range*) between different percentiles) is smallest in CL1 and largest in CL2. The *shape* of GWL time series in terms of both scale (hydrograph magnitude – *sqr.avg*) and slope (e.g. from *rise.avg*) distinguish the clusters quite well from each other. For example, CL3 is particularly noticeable for combining GWL time series with a small slope while CL4 combines such with a large slope.

3.2 Classification of groundwater level dynamics patterns

In the classification task using RF, controls of GWL dynamics were investigated using more than 5,000 unique cluster-attribute relationships. Stable classification accuracy during hyperparameter tuning was achieved with more than 100 trees. Both

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

splitting criteria tested performed equally well in terms of classification accuracy, whereas the consideration of different features for the splits lead to significantly better accuracy (Fig. D1). The final model of this study was set up using the Gini criterion, randomly varied subsets (square root) of features for each split and 700 trees.

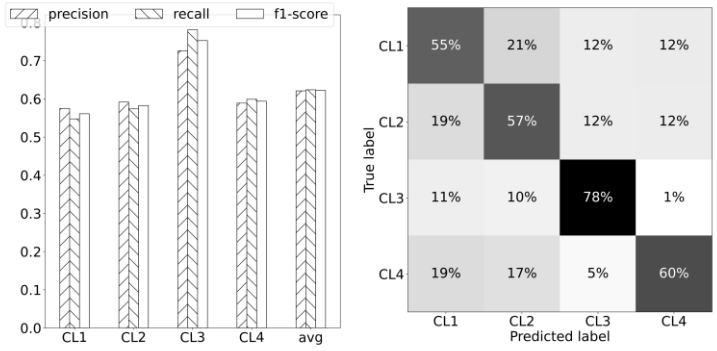


Figure 4 Performance of the classification task using Random Forest (Breiman, 2001). Left: Evaluation metrics of classification (precision, recall, and f1-score) given per cluster and for the weighted average (performance metrics weighted accordingly to the support of each cluster). Right: Normalized confusion matrix. Rows represent the actual cluster, while columns represent the predicted cluster. Thus, the percentages of correctly predicted classes are represented on the diagonal elements and the confusion (i.e. misclassification) is represented on the off-diagonal elements.

With the final model, approximately 62% of the clusters on average were classified correctly, i.e. were assigned to the correct cluster out of four. Comparing the model's result accuracy to a scenario where descriptors were shuffled shows an improvement in accuracy of about 60%. By shuffling the attributes, the relationships between the descriptors and the target variable (clusters) are disrupted, and any observed accuracy in the shuffled model is purely due to chance. These results suggest the presence of moderately strong linkages between the attributes and clusters describing distinct GWL dynamic patterns within the RF model. Fig. 4 shows that the classification performance is distributed slightly unequally across the different clusters. From the confusion matrix, it can be seen that CL3 is most often (78%) predicted correctly and, if not, is either confused with CL1 or CL2, but not with CL4. The other clusters are predicted correctly similarly often, with accuracies ranging from 55% to 60%. Recall and precision of the clusters are quite similar. Considering the slightly different support of each cluster, the classification performs on average similarly well in terms of false positive cluster assignments (e.g. a well site is classified as CL1 when it actually belongs to another cluster), and false negative assignments (e.g. a well site that belongs to CL1 is not classified as CL1 but to another cluster). There is not much difference in classification performance between the different coastal regions.

3.3 Model explanations from SHAP values

SHAP values that were calculated for the classification model are used in the discussion to explain the importance and effects of different controls on GWL dynamics, while general results for the RF model are presented here.

Firstly, by analyzing the correlation of SHAP values, we can detect where feature importance may be affected by information redundancy (Fig. E1). In this study, the redundancy among features tends to be small or moderate, and patterns are not

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

consistent across the clusters. Specifically, the redundancy between feature groups (in correspondence with categories of attributes in Table 1) is minimal for cluster CL3 in the model, whereas it is more evident for cluster CL4. In the case of cluster CL2, there is a redundant contribution of *anthropogenic* features with features derived from *surface waters*, with *climate* features additionally redundant to the prediction of cluster CL1.

Table 2 Qualitative description of the more important and in their effects more differentiable features, starting with the feature group with the greatest importance for the model (Fig. E2), for the prediction of the individual clusters in comparison based on Fig. E3. "<" and ">" mark smaller and larger feature values for one or more clusters compared to one or more other clusters, while there is no unique information for --. *aridity index: see Table 1.

category	attribute/feature (name)	CL1	CL2	CL3	CL4
topography and subsurface depths	elevation	<	<	>	<
	wtd	<	<	>	<
	slope	--	<	>	<
soil and lithology	soil fraction (clay; silt; sand)	<	>	>	<
	aeolian deposits (unconsol_sediment_Ae)	--	<	<	>
	glaciofluviale deposits (unconsol_sediment_Gf)	--	<	>	--
climate	temperature (_min; _mean)	--	>	<	--
	precipitation	>	--	<	--
	aridity*	>	>	<	>
anthropogenic	irrigation	>	--	<	>
	population_density	--	<	<	>
surface waters	distance_coast	<	--	--	>
landuse and landcover	cold temperate moist grassland on plains	>	--	<	>

Secondly, three groups of features are ranked higher in terms of overall feature importance in the global model compared to the remaining three groups (Fig. E2). These are *topography and subsurface depths*, followed by *soil and lithology*, and *climate*, with only minor differences in the importance of the features between the clusters. The largest difference is that the feature group *topography and subsurface depths* are about twice as important for predicting CL3 as the other three clusters, for which the feature groups *soil and lithology* and *climate* are as important as *topography and subsurface depths*. Therefore, features of the group *topography and subsurface depths* primarily play a role in distinguishing CL3 from the other clusters. More in-depth insights into how the model can distinguish the clusters and in particular CL1, CL2, and CL4 from each other through the various features can be derived when using SHAP values to link the feature impacts to their effects (Table 2; Fig. E3). From the individual features, elevation, slope, and water table depth are most important within the feature group *topography and subsurface depths*, soil fractions and the (non)occurrence of specific unconsolidated sediments are most important within the feature group *soil and lithology*, temperature is most important within the feature group *climate*, irrigation is most important within the feature group *anthropogenic*, distance to the coast is most important within the feature group *surface waters* and the (non)occurrence of a specific terrestrial ecosystem type is most important within the feature group *landuse and landcover*.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

Thirdly, prevailing effects that features have on the prediction of clusters in the RF model (Table 2) are those that separate CL3 from the other clusters, effects that specifically separate CL3 from CL4, effects that separate CL4 from CL1 and CL2, and effects that separate CL1 from CL2, although the latter is least successful in the model (largest confusion values in Fig. 4). CL3 is best distinguished from the other clusters by a comparatively large elevation, large water level depth, and large slope. In comparison with CL2, CL3 combines well locations with lower minimum and average temperatures and sediments typical for glaciofluvial deposits. In comparison with CL1, CL3 combines well locations with less precipitation, more coarse-grained soils and less anthropogenic activity. Feature effects between CL3 and CL4 are most often opposite, although less pronounced for climate. Where CL1 is predicted, there are comparatively more coarse-grained soils and rather not sediments typical for aeolian deposits and anthropogenic activity is low. These characteristics from *soil and lithology* as well as anthropogenic activity also separate CL4 from CL2. Less differentiates CL4 from CL1. It is noticeable that the distance of the well to the coast is generally higher in CL4 compared to CL1. The typical well locations in CL1 and CL2 differ in their soil fractions (finer-grained soils in CL1) and anthropogenic activity (larger in CL1).

Lastly, SHAP values for individual wells all located in the same region can be used to understand to what extent the results of the feature importance and effects also apply regionally. SHAP values were analyzed for four well sites located in northern Germany (Fig. 5) with the highest model's probability for predicting each cluster in the test dataset (Fig. E4-E7). While the wells most likely predicted as CL1 and CL3 belong to these clusters, the wells most likely predicted as CL2 and CL4 belong to CL3. That CL1 is correctly predicted is due to low elevation and wtd at the well location as well as the presence of grassland and mean temperatures around 10 °C. With these characteristics, the model predicts CL1 and CL3 over CL2 and CL4 in the case study region. CL3 is also correctly predicted because at the well location both elevation and anthropogenic activity are high and incorrectly predicted at the other well location as CL2 when the opposite is true. As CL4, CL3 is also incorrectly predicted when elevation is low and the clay content is high. The described correlations are in line or not contradictory to the global correlations of Table 2.

4. Discussion

4.1 Hydrogeological similarity and scaling effect

Four predominant types of GWL dynamics were identified for coastal wells distributed across five continents. Our results show, that shallow aquifers exhibit more diverse GWL dynamics than deeper aquifers. This suggests that absolute GWL depth is less determinant in differentiating global GWL dynamics than might be assumed based on the clustering of direct time series from a single particular region, as in Rinderer et al. (2019) where in particular mean water level and amplitude of variation determine clusters of GWL dynamics. Based on the spatial distribution of clusters, wells exhibit distinct regional structures that strongly imply the presence of underlying spatial controls, e.g. CL3 type GWL dynamics in northern Germany tend to coincide with Geest land, while CL1 type GWL dynamics are more often found in the marshlands (Fig. 5a). Another example

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

is that the topographical elevated Veluwe region in the center of the Netherlands is the only region at the North Sea where many wells are assigned to CL3 (Fig. 1). However, GWL dynamics of different types are also frequently found in wells within a very short distance from each other or even on top of each other (in the case of multiple groundwater storeys). The first is more surprising than the second because it suggests a small-scale separation of groundwater landscapes. In line with the findings by Wunsch et al. (2021), this result shows that nearby wells do not necessarily have a higher degree of similarity in GWL dynamics compared to more distant wells. Regionally similar, but locally diverse GWL dynamics highlight the complexity of generalizing GWL dynamics on different spatial scales. Yet, the fact that our results include spatial patterns at local, regional, and global scales (considering the globally distributed clusters of GWL dynamic types), speaks for their robustness with respect to scale dependency. Before, studies on GWL dynamics have resulted in system understanding at local and regional scales (Wunsch et al., 2021; Giese et al., 2020; Haaf et al., 2020). Although this may be more straightforward for applied groundwater management and monitoring at these scales, this study is the first to provide insights into generalizations of GWL dynamics using a global dataset.

The methodological approach underlying the GWL dynamics analysis offers much potential but also has limitations and uncertainties with respect to the validity of the generalizations generated. We found that using indices for analyzing GWL dynamics is efficient, and offers comparability and interpretability. In these terms, this approach might be superior compared to directly analyzing GWL time series or only focusing on the long-term mean or trends. Heterogeneous time series with different seasons, time series lengths, and periods can be combined in this methodological approach with robust results as previous studies have shown (Heudorfer et al., 2019; Wunsch et al., 2021), thereby reducing the dimension of the time series and potentially better establishing cause-and-effect relationships. From the clustering results, we can see that indices successfully separate between climatic forcing (as present in seasonality) and physiographic forcing, while in practice it remains challenging to identify what drives GWL dynamics at various scales and settings (Blumstock et al., 2016; Moeck et al., 2020) and to distinguish between climatic and other natural factors versus anthropogenic impacts on GWL dynamics and trends (Wriedt, 2017; Lischeid et al., 2021). Furthermore, from a global perspective, GWL time series are rarely sufficiently long to recognize the dimension of climate change. Although indices are also affected by the length of the GWL time series (Heudorfer et al., 2019), they can help identify time series with "high-amplitude, low-frequency variability," such as in the Normandy (Paris Basin, France; compare Baulon et al., 2022b). Groundwater well sites with dominant interannual variations from multi-annual to decadal scales identified by Baulon et al. (2022b) in this area match well sites with CL3 type GWL dynamics found in this study (Fig. C6). At these well sites, the likelihood of approaching short-term groundwater drought is greater if the prevailing low-frequency variability is associated with a downward phase, while the annual groundwater level status and last winter's groundwater recharge have less influence on short-term drought (Baulon et al., 2022b).

Despite that general robustness of indices was found starting at time series with a minimum length of four years (Heudorfer et al., 2019), uncertainties remain regarding the representativeness of the generalizations for individual wells with this time series period length, but less so for the generalizations themselves. Furthermore, Heudorfer et al. (2019) originally defined some of the indices for time series with daily resolution, and decadal GWL dynamics are only indirectly represented by indices in our

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

study (e.g. seasonality in the annual hydrograph). Therefore, further investigation of the effects of different temporal resolutions and time series lengths, as performed in Papacharalampous et al. (2023) for hydroclimatic time series, would be of interest. High-resolution time series are, for example, required for analyzing the interaction of groundwater with the sea (Haehnel et al., 2023). Accounting for SWI in GWL dynamics pattern analysis is best supported by pattern recognition or correction with groundwater chemistry data (Narvaez-Montoya et al., 2023; Rau et al., 2020). However, the lack of GWL time series with high temporal resolution and long time series length currently hinders the expansion of such studies. Furthermore, with respect to the groundwater systems and processes represented by the GWL dynamics patterns, it should be emphasized that there may be a bias in favor of the typical GWL dynamics in northwestern Europe because most of the wells in the dataset are located there. Clustering analysis has been instrumental in capturing the GWL dynamics patterns observed in this study. The utilization of the k-means algorithm has enabled us to represent these patterns at a high level, providing valuable generalizations. However, if in other research questions the necessary level of pattern recognition is smaller, clustering algorithms that do not require both clearly distributed data and clearly defined clusters have the potential to identify additional clusters and uncover more nuanced dynamics within the dataset.

4.2 Cause-and-effect relationships

We analyzed whether the identified types of GWL dynamics are meaningful, i.e., consistent with expected patterns (Yang and Chui, 2021), although machine learning can also reveal undetected and previously unexplained patterns. We generally expect patterns as a result of the multitude of natural and anthropogenic factors influencing GWL dynamics. Meaningful GWL dynamics further hold potential to derive cause-and-effect relationships between GWL dynamics and their driving forces, that are comparatively rare yet. Thus, in this study, driving forces were linked to GWL dynamics patterns in an RF classification task. Here, many different environmental attributes describing surface and subsurface processes, potentially associated with groundwater recharge and discharge, were used to explain (dis)similarities within and between the identified patterns. While comparing individual aspects of the hydrograph with potential controlling factors can only explain specific aspects of GWL dynamics (e.g. Haaf et al. (2020) related flashiness (BFI) to focused recharge as a consequence of depressions or connectivity to streams), the holistic analysis of various GWL dynamics aspects against controls enables the estimation of the influence of multiple controls on processes defining groundwater quantity over time. Since GWL dynamics in coastal regions can be similar in different regions globally, similar GWL dynamics are not necessarily the result of the same cause-and-effect relationships. Rather, it can be expected that there are complex interlinkages of multiple controls that favour hydrogeological similarity at different manifestations of the individual attributes. We therefore confirm our initial expectation that similar GWL dynamics derived from indices are associated with multiple processes and, vice versa, that the important controls and processes can be identified and distinguished by a combination of indices. This is in line with previous findings with hydro-climatic time series (McMillan, 2020; Beck et al., 2016).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

The results have shown, that relationships between GWL dynamics patterns and controls can be generalized for the global scale to some extent. Notably, the importance of different attribute categories (Table 1) for shaping GWL dynamics is related to hierarchical relationships where climate, topography, and soil conditions have a greater influence on the location and characteristics of surface waters and landuse and landcover than the other way around (Fig. E2). This implies that the direct influencing factors offer more explanatory power. Then again, direct influencing factors in the attribute dataset are more often based on actual observations. From the importance of elevation and water table depth, especially for the determination of CL3 type dynamics and differentiation from other dynamics that are more often found in near-surface aquifers, it can be concluded that these attributes mainly describe the increasing damping effect of groundwater recharge with increasingly long soil passages. The generally high importance of climatic attributes for GWL dynamics in this study - either only in the long term or also in short periods - is scientific consensus making climate change usually part of assessments of changes in groundwater systems (Riedel and Weber, 2020). Also, within the group of climatic attributes, the importance of direct attributes such as temperature outweighs derived attributes such as the aridity index. According to Table 2, certain types of unconsolidated sediments and a terrestrial ecosystem can distinctly indicate or exclude specific GWL dynamics types. These manifestations of categorical attributes are likely to serve as proxies for the general importance of their class. Due to the multitude of manifestations, only the most common ones likely appear in the importance ranking in Fig. E3 (e.g. aeolian deposits mainly exist in the Netherlands which also provides the most wells in the dataset).

From the comparably minor importance of anthropogenic attributes for GWL dynamics, it can be assumed that global groundwater quantity is less influenced by anthropogenic than by natural controls; but this was also the expectation with using data derived from monitoring networks and after quality control. However, anthropogenic influence is difficult to disentangle from natural controls also in this study. It is unclear how much of the explanatory content of the anthropogenic attributes is based on anthropogenic activities and how much is due to correlations with natural controls (Fig. E1). Such interactions could be that irrigation is higher in dry climates and that there are more sealed surfaces where population density is high. Anthropogenic activities such as irrigation and groundwater pumping typically have a localized effect. They have the potential to overprint different types of GWL dynamics even when the hydrogeological setting is the same (Sorensen et al., 2021). Together with small-scale lithological and hydrogeological peculiarities, they likely contribute the most to the GWL dynamic patterns on a small scale.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

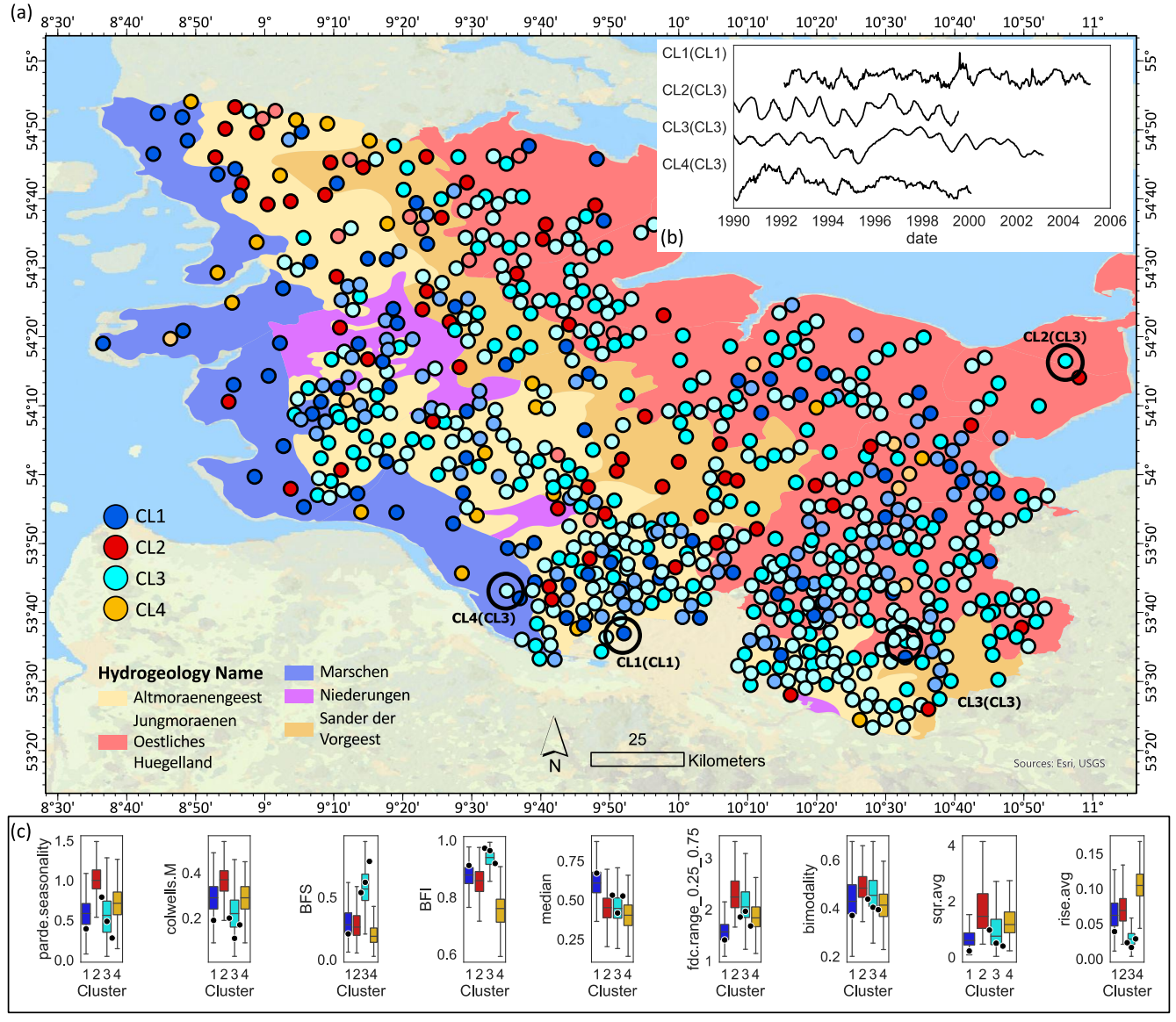


Figure 5 GWL dynamics patterns in the German federal state Schleswig-Holstein. (a) Cluster assignment to well locations and hydrogeology map of the German federal state Schleswig-Holstein (LfU-SH, 2003). Overlapping well markers are jittered at a minimum spacing of 500 meters and thus no longer represent the original well locations. Lighter cluster colors mark lower aquifer storeys. The cluster number in parentheses marks the true cluster from k-means clustering. (b) Exemplary time series of each cluster from the encircled well sites in (a). (c) Within-cluster variability of the example indices that each represent a component of the Heudorfer et al. 2019 typology from Fig. 3 plotted with indices values (dots) of the encircled well sites in (a) where the dots between CL2 and CL3 and CL4 and CL3 mark the wells that have the highest probability for CL2 and CL4 in the RF model but belong to CL3 according to the indices (see Fig. C5 for within-cluster variability for all indices that are plotted together with indices values of the selected well sites).

Moderate accuracy of the RF model means that parts of the GWL dynamics derived from indices are not well predicted based on the used attributes. Hence, machine learning tools that are trained to find cause-and-effect relationships face the same

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences
Discussions

challenge of process-based groundwater models: limited availability and quality of explanatory data, especially subsurface information. Even if there is great progress in the availability and quality of geoscientific data, surface attributes derived for the global scale are still biased by uncertainty in the data itself (Peng et al., 2017), coarse spatial resolution (e.g. here the minor explanatory power of attributes from surface waters might also be related to a coarse spatial resolution of about 55 km) and by the period they cover. In the presented approach no dynamic attributes were used because they are not available for all attribute categories at the global scale. For instance, land use has a large influence on the water balance with land use conversion (Mishra et al., 2014). That explanatory information is missing in the RF model for differentiating GWL dynamics at the local to regional scale can be observed, for example, in the case study region in Northern Germany (Fig. 5), where the temperature should not vary significantly between well locations. However, according to SHAP values, temperature is the most important following elevation and water table depth (WTD) for the prediction of GWL dynamics types at specific wells (Fig. E4-E7).

4.3 Case study

Finally, the question also arises to what extent it is possible to explain the generalized GWL dynamics derived with the indices approach of this study also regionally if attributes were only good enough. A ground-truthing of the GWL dynamics patterns using expert knowledge from the case study region in northern Germany will shed light on this in the following. This case study was selected because a hydrogeological map and information on the aquifer storey of the wells included in the GWL time series dataset are available, and at the same time, the well density is high, which allows the study of spatial patterns. To a large extent, the near-surface aquifer in this area is a pore aquifer of silicate type formed during the Ice Age (LfU-SH, 2003; Otto, 2001). The hydrogeological situation of the shallow groundwater can broadly be divided from west to east into three major landscapes: (a) low-lying marshlands (*Marschen*), (b) slightly raised Geest landscapes (*Altmoraenengeest*), the lower outwash plain (*Sander der Vorgeest*) and intervening depression zones comprising the river valleys (*Niederungen*) and (c) morainic uplands with strong relief (*Jungmoraenen Oestliches Huegelland*). Fig. 5a shows that the GWL dynamics in the marshlands predominantly belong to CL1 (59 %), especially in the upper two aquifer storeys (68 %) and close to the sea, followed by CL3 (20 %) which combines more GWL dynamics of the lower aquifer storeys (60 %) and well sites at the border to the Geest. The Geest and uplands are characterized by GWL dynamics that mainly belong to CL3, with some exceptions including the Geest landscape in the north close to the Danish border (GWL dynamics are of mixed type), the area around the major depression in the north (GWL dynamics mainly belong to CL1) and the eastern uplands at the southernmost bay (GWL dynamics mainly belong to CL1). In the lower Geest, GWL dynamics belong more often to CL2 (23 %) than in other landscapes. As in the marshlands, type CL3 GWL dynamics are more often associated with lower aquifer storeys compared to the other clusters in these areas (63-78 %), with the difference that CL3 is also present to a similar degree in the upper aquifer storeys in the Geest and uplands (45-67 %). Fig. 5b displays example time series of wells that the RF model identifies as having the highest probability for each cluster. The cluster in parentheses indicates the actual cluster based on the indices values shown in Fig. 5c. In cases where CL3 was mistaken for CL2 and CL4, the example wells mostly exhibited indices

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences Discussions

within the CL3 distribution (Fig. 5c). While there are minimal visual differences in the time series between CL2 and CL3 at the well location where CL2 is predicted, some distinctions can be observed in Fig. 5b when CL3 is misidentified as CL4. The inability of the RF model to correctly predict these instances can therefore mainly be attributed to two factors. Firstly, the GWL dynamics at the predicted wells for CL2 and CL4 reside closer to the border between different dynamic types, making them more challenging to differentiate. Secondly, the lack of explanatory potential provided by the attributes can be assumed to also contribute to the model's failure: Expert information suggests that small-scale differences in GWL dynamics in the case study region can be explained by lower aquifer storeys with good hydraulic connections to upper storeys and surface waters. The presence of both CL1 and CL3 dynamics in the higher Geest and upland regions, regardless of groundwater storey, can be attributed to differences in the soil covering, water table depth, and the existence of both unconfined and confined upper aquifers in these landscapes (LfU-SH, 2003). However, the RF model offers some global explanations that apply to this regional context as well. For instance, in the marshlands and the northern depression zone, where impermeable sediments of small to medium thickness overlay the aquifer (LfU-SH, 2003), the prevalence of CL1 and CL4 type wells is higher (Table 2). We have demonstrated some physical explainability and plausibility of index-derived GWL dynamics across different scales. However, the main limitation of this study's approach is the limited explanatory power of the available attributes describing potential controls of GWL dynamics. Attributes derived solely from natural and anthropogenic characteristics at the surface alone may not fully capture the salient controls, particularly considering the low resolutions currently available for large spatial scales and that there is still no best way of identifying groundwater catchment areas at the surface without detailed information provided by bore profiles. One way to deal with this limitation is by utilizing attributes that directly represent influencing factors or attributes closer to actual observations in combination with machine learning techniques. At present, disentangling the influence of anthropogenic activities from the natural controls of GWL dynamics and accurately predicting the impact of climate change on GWLs without incorrectly accounting for water withdrawals remains a challenge in the field of machine learning, which is of great importance for sustainable groundwater management.

5. Conclusions

In summary, this study provides new insights into the hydrogeological behavior of groundwater in coastal regions. With the presented methodology, the integration of both local variations and global patterns shed light on how similarities in GWL dynamics can be linked to the impacts and effects of various controlling factors of groundwater recharge and discharge processes. We demonstrated that the proposed methodology allows capturing the GWL dynamics patterns at different granularity for high-level generalizations while retaining the potential to uncover more nuanced dynamics-controls relationships at smaller scales within the dataset.

Our results indicate that hydrogeological similarity manifests globally through the identification of four clusters representing distinct patterns of GWL dynamics. Similar patterns are observed across different environments and climates globally, while specific cluster compositions vary among regions, suggesting complex interlinkages of controlling factors. Therefore, it is

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023


Hydrology and Earth System Sciences Discussions

important to consider GWL dynamics patterns at multiple scales as extrapolating GWL dynamics from a single well to the surrounding area may not be valid. General patterns and similarities in GWL dynamics identified for coastal aquifers at the global scale can assist in developing broader frameworks for groundwater management. Overall, dynamics-control relationships are dominated by (a) topography and groundwater depth mainly determining the responsiveness of groundwater systems that well separates a single cluster from three others, and (b) climate and soil characteristics that differentiate these three clusters with high short-term and interannual variability. Furthermore, this study emphasizes the challenges posed by limited, especially subsurface, data in capturing the complexity of GWL dynamics. While hydrogeological similarity offers potential for transferring GWL dynamics from monitored to unmonitored sites, planning requires more accurately explained GWL dynamics by controlling factors than is currently possible with available attribute data from the global scale. We suggest relying more on groundwater hydrographs and derived indices for understanding GWL processes when attribute data is scarce. Additionally, machine learning techniques combined with hydrograph information show promise for improving our understanding and predictive capabilities in addressing GWL complexities.

Appendix A: Groundwater data

Table A1 List (country-wise alphabetical) of governmental agencies or portals from which groundwater data were obtained.

number	governmental agency/portal	country	access year
1	Australian Groundwater Explorer, Bureau of Meteorology (http://www.bom.gov.au/water/groundwater/explorer)	Australia	2021
2	DOV, Datenbank Ondergrond Vlaanderen (https://www.dov.vlaanderen.be)	Belgium	2020
3	RIMAS, Integrated Groundwater Monitoring Database, Servicio Geologico de Brazil (http://rimasweb.cprm.gov.br/layout/pesquisa_complexa.php?rimas=true)	Brazil	2022
4	Department of Environment and Climate Change, Nova Scotia (https://beta.novascotia.ca)	Canada	2021
5	GIN, Groundwater Information Network, Canada (http://gw-info.net)	Canada	2021
6	Ministry of Environment and Climate Change Strategy, British Columbia (https://www2.gov.bc.ca/gov)	Canada	2021
7	GEUS, Geological Survey of Denmark and Greenland (https://eng.geus.dk)	Denmark	2019
8	ADES, Portail national d'accès aux données sur les eaux souterraines, France (https://ades.eaufrance.fr)	France	2020
9	LLUR, Landesamt für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (https://www.schleswig-holstein.de/DE/landesregierung/ministerien-behoerden/LLUR/llur_node.html)	Germany	2020
10	NLWKN, Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (https://www.nlwkn.niedersachsen.de)	Germany	2021
11	Staatliches Amt für Landwirtschaft und Umwelt Mittleres Mecklenburg (https://www.stalu-mv.de/mm)	Germany	2020
12	Staatliches Amt für Landwirtschaft und Umwelt Vorpommern (https://www.stalu-mv.de/vp)	Germany	2021
13	Environmental Protection Agency, Ireland (https://www.epa.ie)	Ireland	2020
14	Hydstra, Department of Water and Sanitation, South Africa	South Africa	2021
15	Geological Survey of Sweden (https://www.sgu.se)	Sweden	2021
16	DINOloket, Geological Survey of the Netherlands (https://www.dinoloket.nl)	The Netherlands	2020

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

17	California Department of Water Resources (https://wdl.water.ca.gov)	United States of America	2022
18	NWIS, National Water Information System, United States Geological Survey (https://waterdata.usgs.gov/nwis/gw)	United States of America	2020

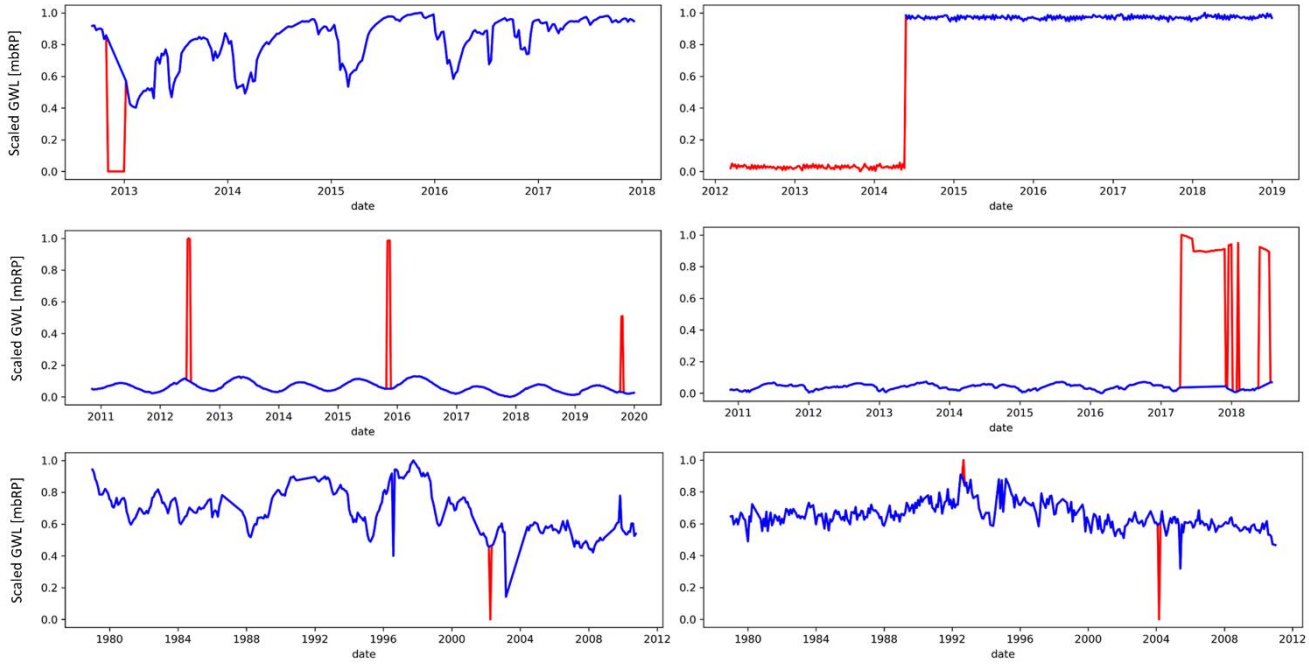


Figure A1 Examples of time series outlier removal with DBSCAN (scikit-learn library in Python): Raw (scaled) time series are shown in blue, and the anomalies identified by DBSCAN are shown in red. The use of DBSCAN is less successful or difficult to verify in the bottom two examples (left: The first abrupt change in direction of the GWL between 1996 and 2000 is not identified as an outlier, although one would expect one based on the magnitude and position of the change; right: From visual inspection, it is unclear whether the identified outliers in both directions are due to measurement error, human activity or reflect natural behavior of the groundwater system e.g. to extreme events).

Appendix B: Indices

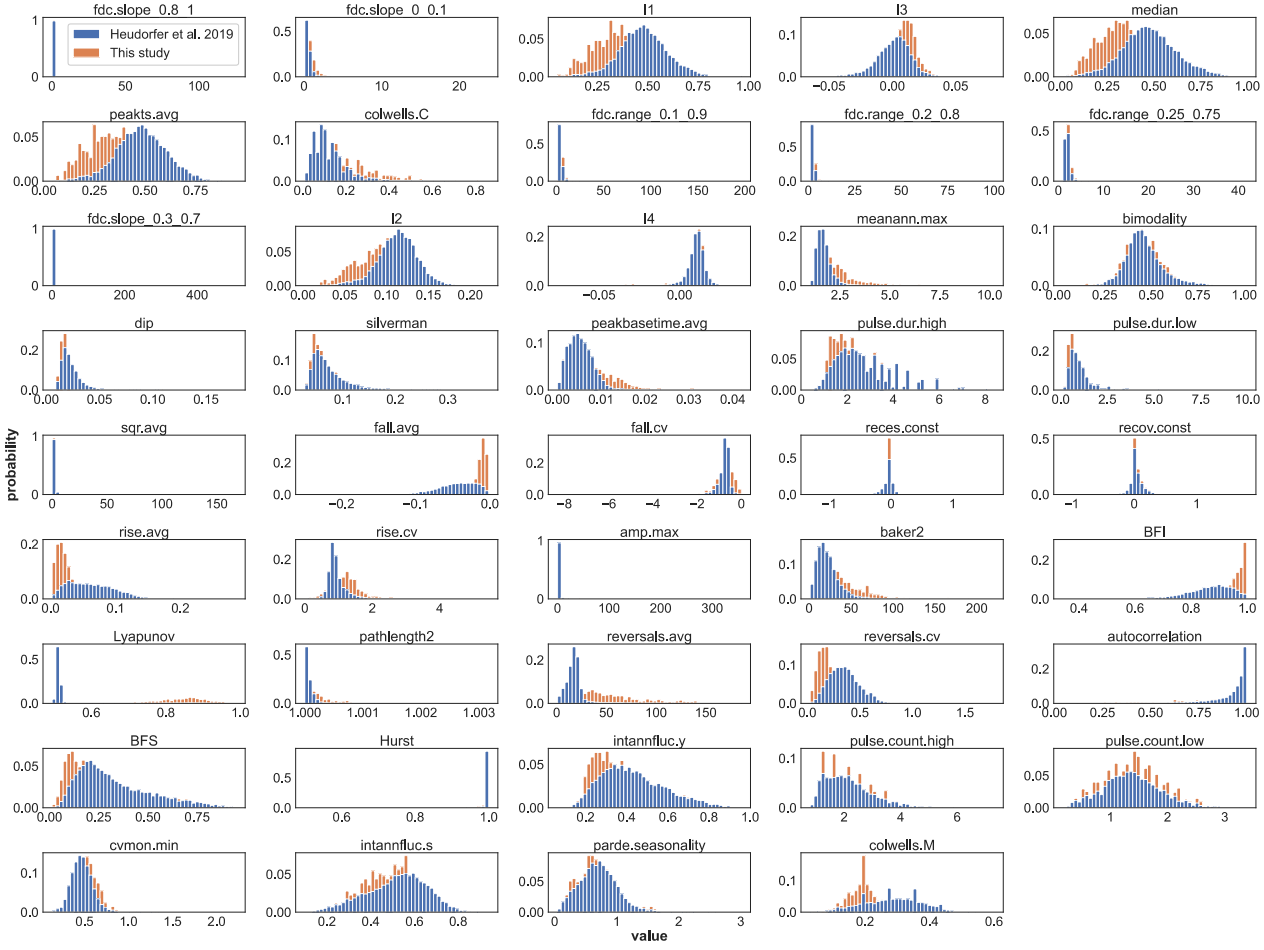
Table B1 List of GWL dynamics typology and indices used in this study. Differing names in brackets refer to deviating abbreviations in Heudorfer et al. (2019).

aspect	component	indices names
distribution	boundness	fdc.slope_0.8_1 (dc.slp.u); fdc.slope_0_0.1 (dc.slp.l); l1; l3; median (med); peakts.avg
distribution	density	colwells.C (colwell.C); fdc.range_0.1_0.9 (dc.rng.01.09); fdc.range_0.2_0.8 (dc.rng.02.08); fdc.range_0.25_0.75 (dc.rng.025.075); fdc.slope_0.3_0.7 (dc.slp.m); l2; l4; meanann.max (avg.ann.max)
distribution	modality	bimodality (bimod); dip; silverman (bandwd)

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

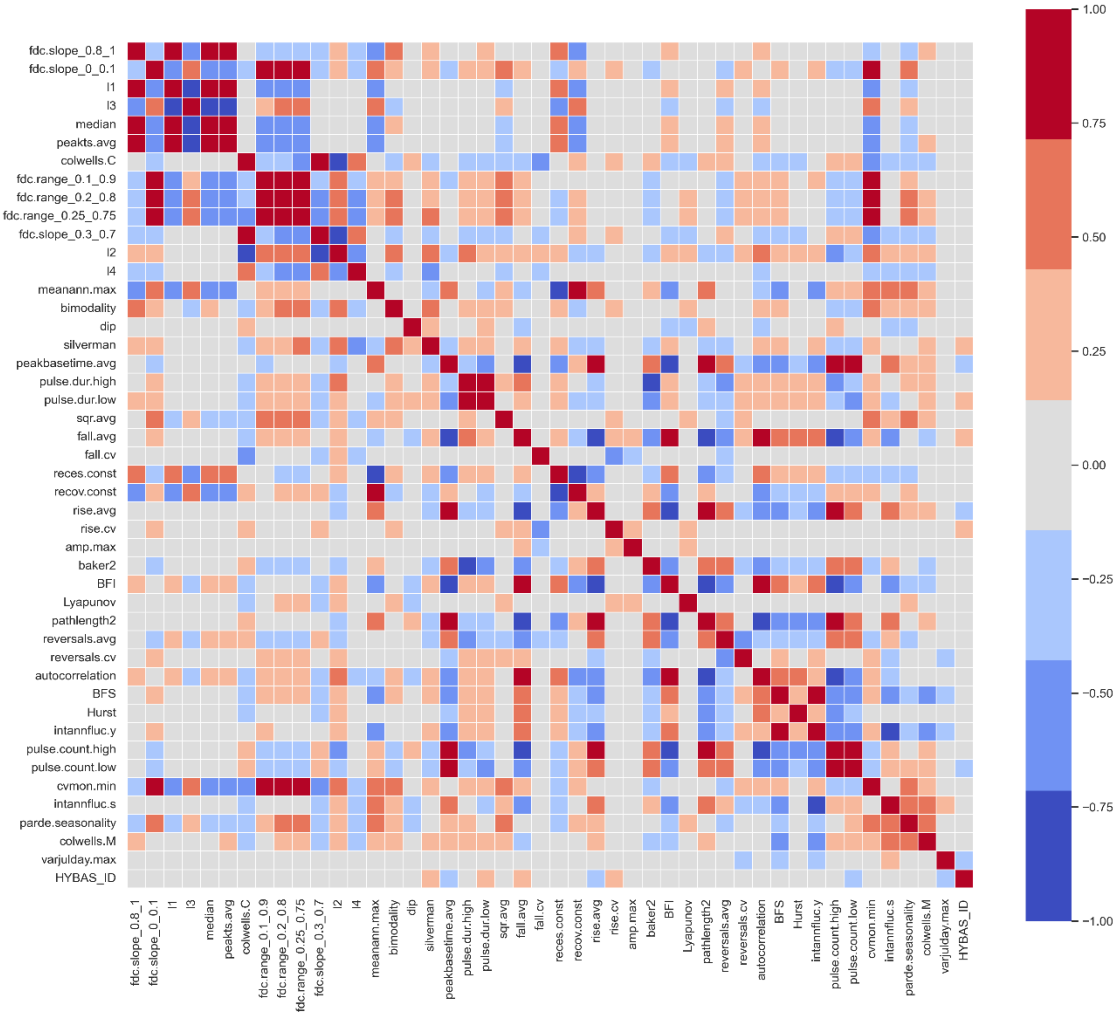
shape	scale	peakbasetime.avg (peakbase.avg); pulse.dur.high (pulse.dur.h); pulse.dur.low (pulse.dur.l); sqr.avg
shape	slope	fall.avg; fall.cv; reces.const; recov.const; rise.avg; rise.cv
structure	amplitude	amp.max
structure	flashiness	baker2 (baker); BFI; Lyapunov; pathlength2; reversals.avg (rev.avg); reversals.cv (rev.cv)
structure	interannual variation	autocorrelation (autocor); BFS; Hurst; intannfluc.y (iaf.y); pulse.count.high (pulse.count.h); pulse.count.low (pulse.count.l)
structure	seasonality-magnitude	cvmon.min; intannfluc.s (iaf.s); parde.seasonality (parde)
structure	seasonality-timing	colwells.M; varjulday.max (newly introduced instead of varjuld.min: CV of date of annual maximum head)



https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

555 **Figure B1 Comparison of the value distributions of the indices calculated in this study (Table B1) with the values published by Heudorfer et al. (2019) (data set: Haaf and Heudorfer, 2018). The magnitude and variability of values for many indices are similar. Shifted or significantly different ranges of values are observed for some indices. Such differences exist for indices of different aspects of the hydrograph (structure, distribution, and shape).**



560 **Figure B2 Correlation matrix of the index values calculated in this study.**

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

Appendix C: Clustering

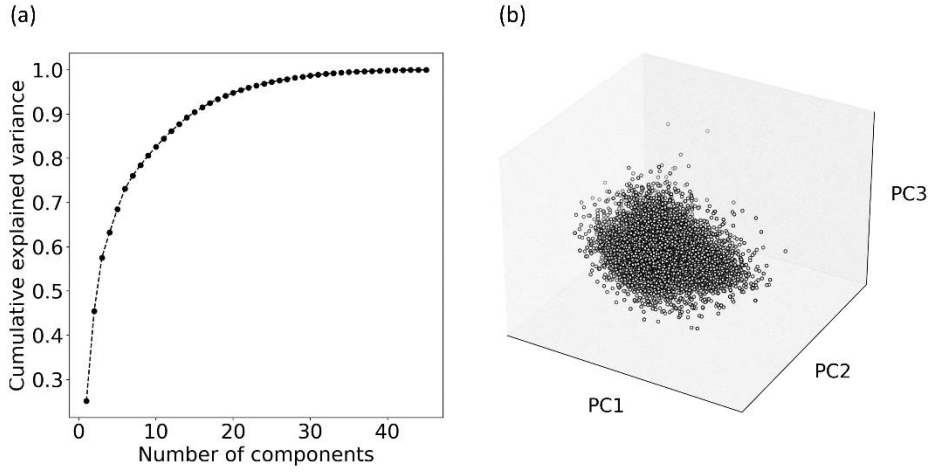


Figure C1 Principal Component Analysis (PCA) with the indices dataset. (a) Cumulative explained variance with an increasing number of Principal Components (PCs). (b) Dataset value space of the first three PCs (which together explain more than 50% of the variance of the dataset).

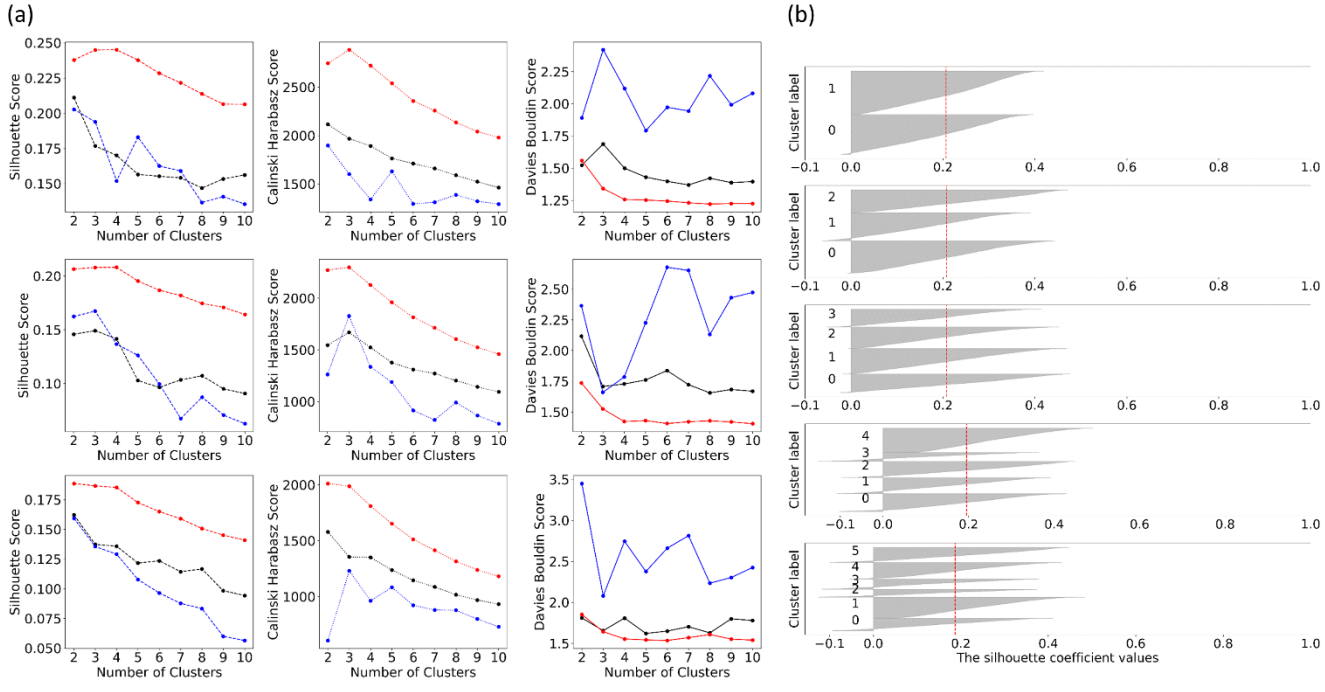


Figure C2 Evaluation metrics – Silhouette (Rousseeuw, 1987), Calinski-Harabasz (Caliński and Harabasz, 1974) and Davies-Bouldin (Davies and Bouldin, 1979) – used to find the best cluster separation. The Silhouette score ranges from -1 to 1, where a higher value

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

indicates better-defined clusters. Better clustering results are also indicated by higher values for the Calinski-Harabasz score and lower values for the Davies-Bouldin score. (a) Metrics are shown for different proportions of explained variance represented by Principal Components (PCs) (top to bottom line: 60%, 70%, and 80%) and for different clustering algorithms (k-means (red), Gaussian mixture (blue) and hierarchical clustering (black)). (b) Silhouette scores for various clusters using the k-means algorithm and PCs representing 70% of explained variance.

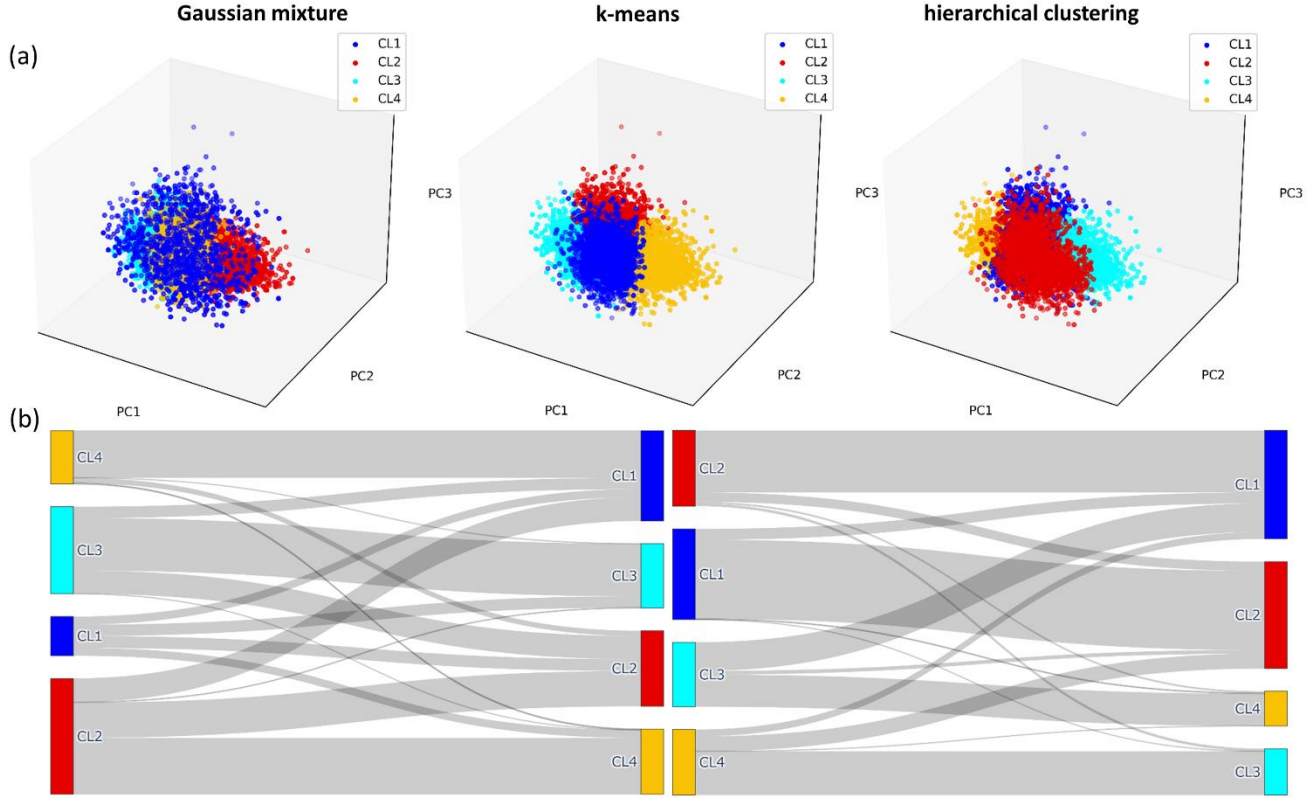


Figure C3 Assignment of the time series to four different clusters by three algorithms (Gaussian mixture, k-means, and hierarchical clustering). (a) Illustrative cluster assignment within the range of values of the first three principal components (PCs) of the dataset. (b) Sankey plot allows comparison of the quantitative distribution of the samples among the clusters (left between Gaussian mixture and k-means and right between k-means and hierarchical clustering). The majority of samples are assigned to the same cluster for k-means and hierarchical clustering (shown in different order).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

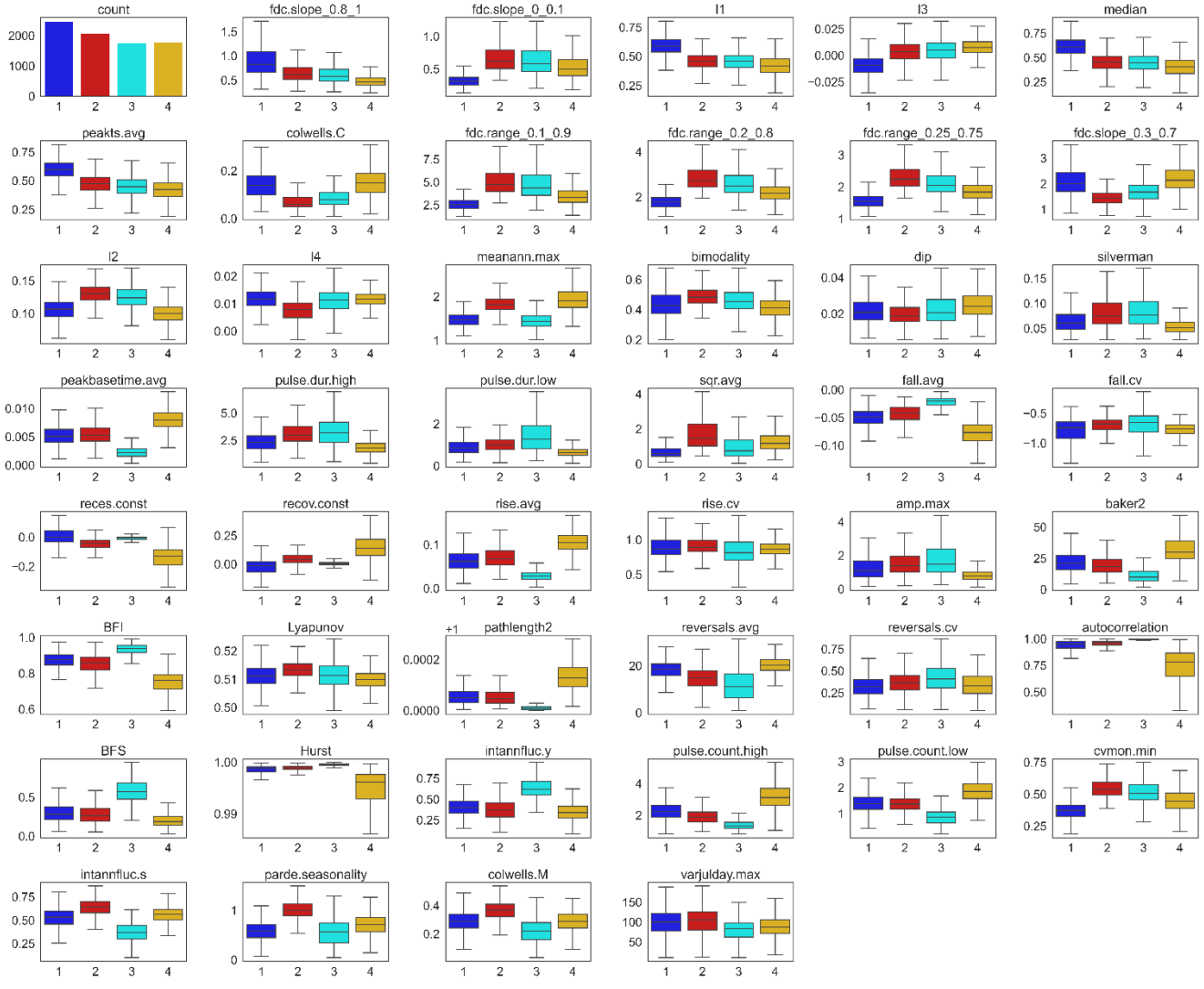


Figure C4 Within-cluster variability of all indices used for clustering.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

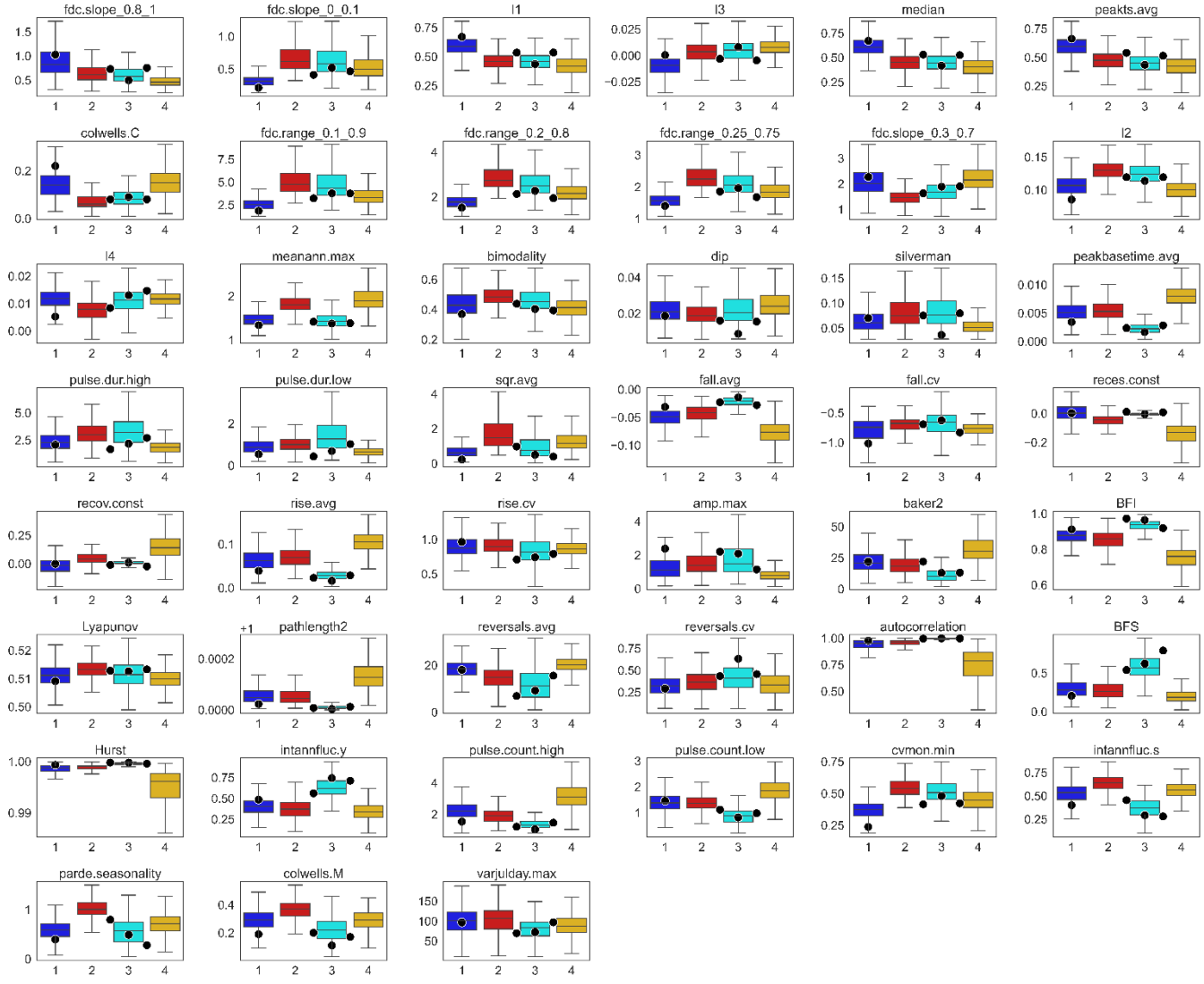


Figure C5 Within-cluster variability of all indices used for clustering (Fig. C4) plotted with indices values (dots) of the four example wells from the case study region in northern Germany (Fig. 5a). The points drawn between CL2 and CL3 and CL4 and CL3 are wells with the highest probability for CL2 and CL4 in the Random Forest model but belong to CL3 according to the indices.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

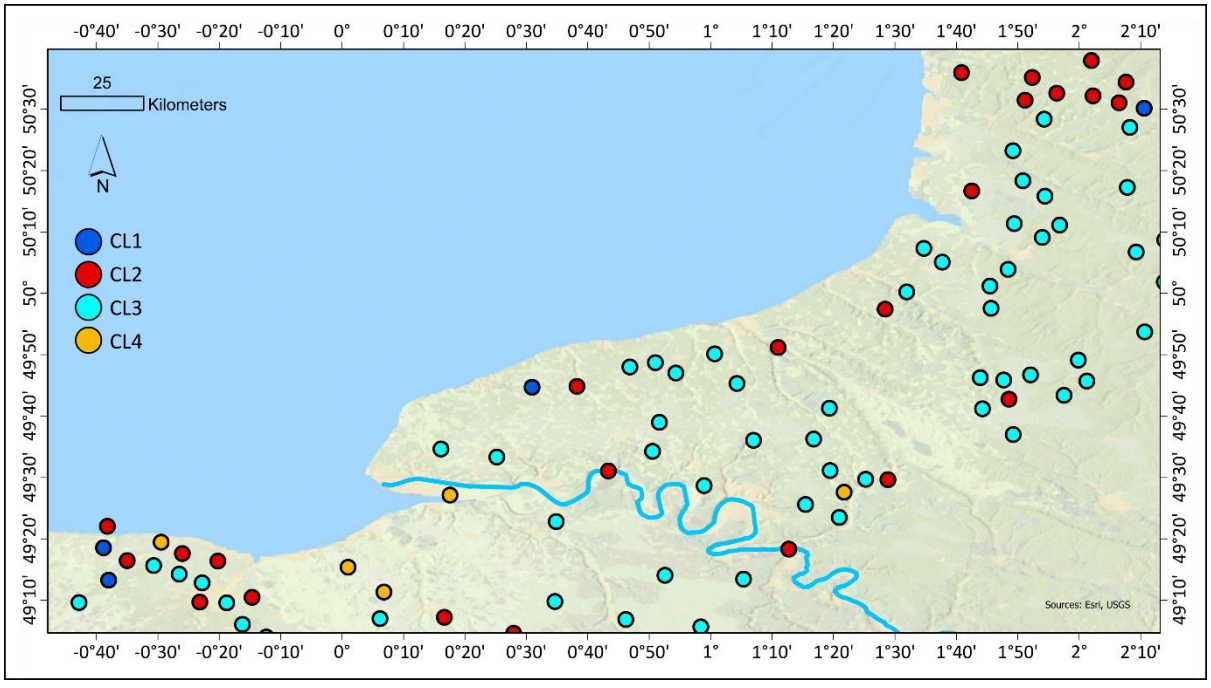
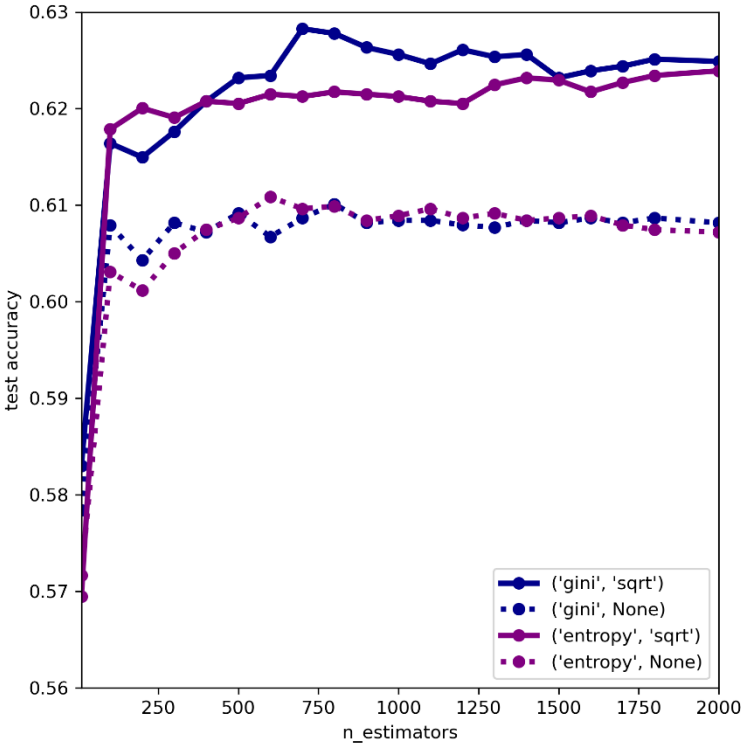


Figure C6 Assignment of clusters to well locations in the Normandy, Paris Basin displayed with the Seine river. Overlapping well markers were jittered at a minimum spacing of 500 meters and thus no longer represent the original well locations.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

Appendix D: Random Forest classification



590

Figure D1 Random Forest hyperparameter tuning within a 5-fold cross-validation framework using the training data.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions

Appendix E: SHAP

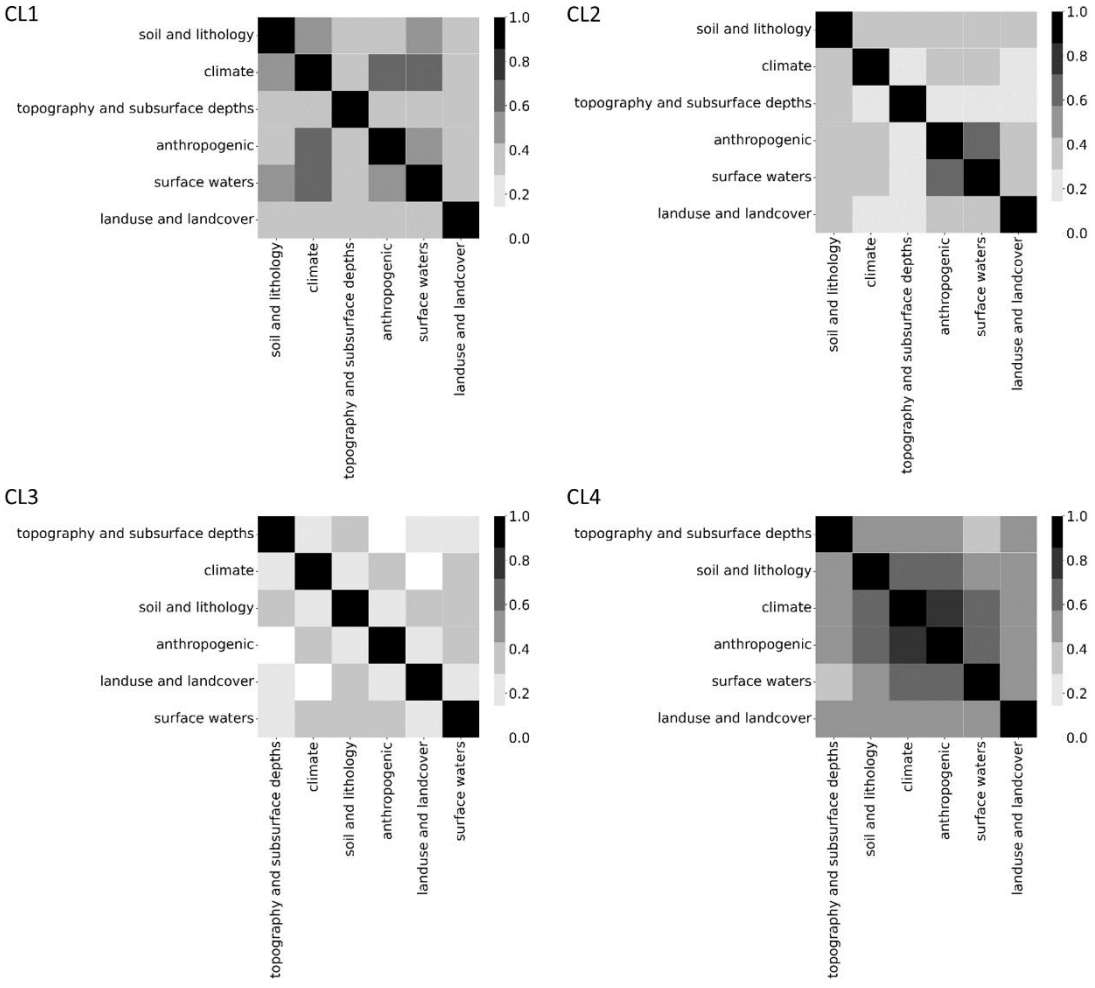


Figure E1 Absolute spearman correlation of SHAP values for aggregated features (SHAP values added up for features from the same attribute categories from Table 1).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

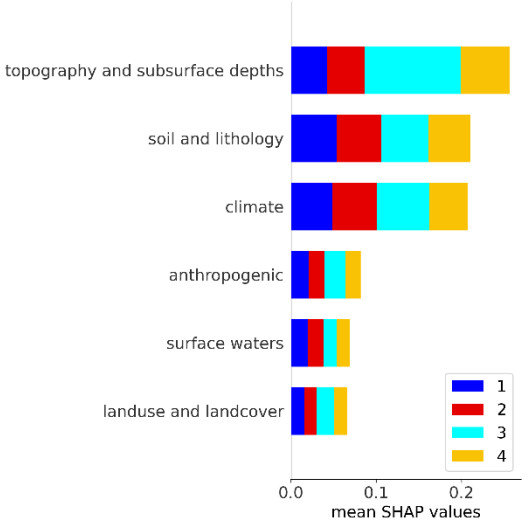
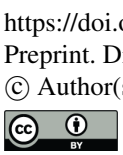


Figure E2 Overall SHAP feature importance stacked for individual clusters for aggregated features (SHAP values added up for features from the same attribute categories from Table 1).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

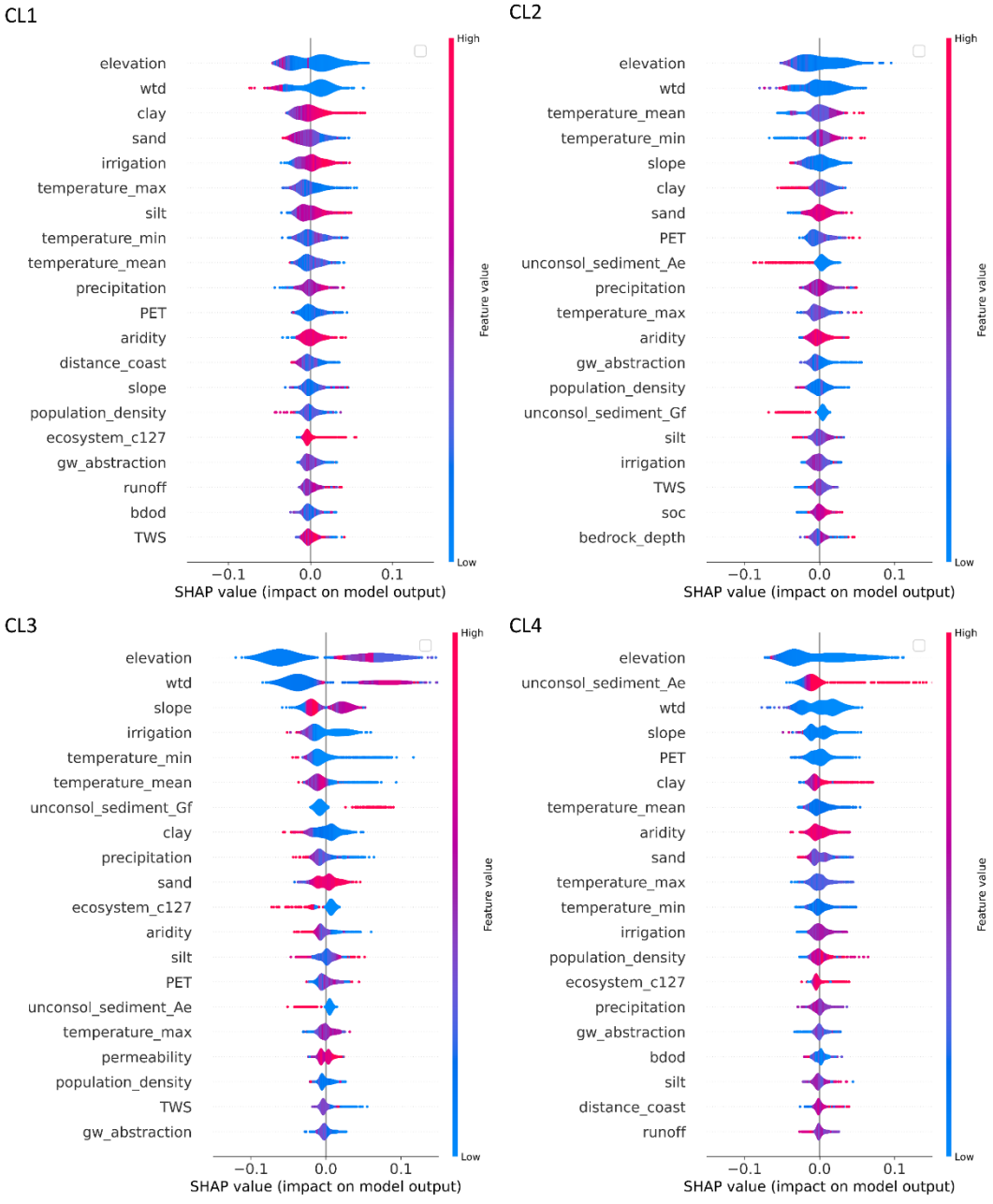


Figure E3 Violin plots show the distribution of SHAP feature importance and feature effect for the twenty highest-ranked features of each cluster. SHAP values describe the impact of a given feature on the prediction of the model for a given cluster (the prediction of "True" for the selected cluster). A red color corresponds to high feature values and a blue color corresponds to low feature values (i.e., for one-hot-encoded features, a value of 0 (blue) corresponds to the presence of the feature class and a value of 1 (red) corresponds to the absence of the feature class). Overlapping instances widen the violin shape in the direction of the y-axis and the violin at that position is colored according to the average feature value. SHAP values were not aggregated for this figure.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

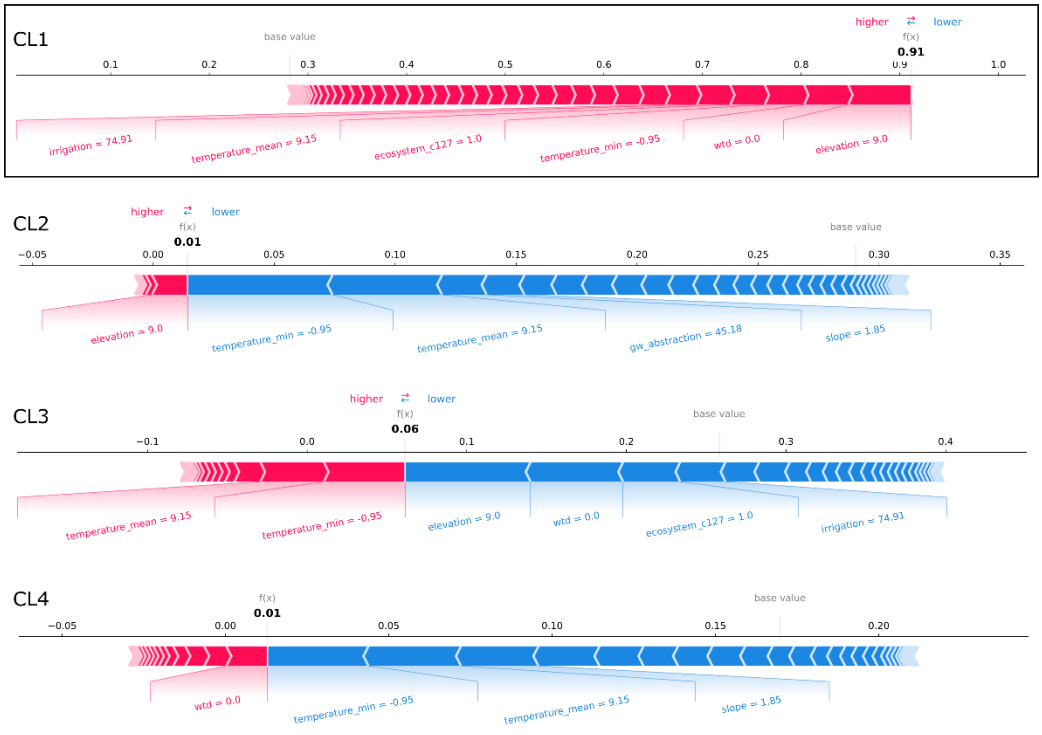


Figure E4 SHAP force plots explaining the correct prediction of the CL1 exemplary well site from the case study region in northern Germany (Fig. 5a) (from top to bottom: CL1, CL2, CL3, and CL4). At this well site, the probability that CL1 is predicted is the largest in the case study (test dataset). The average scores of all classifications made by the Random Forest model for the training dataset are 0.28, 0.29, 0.26, and 0.17 respectively (base values). The model's probability scores for predicting the clusters at the exemplary well site (bold printed numbers) sum up to 100 %. Features that are important for the respective predictions are displayed with their values. Their importance and effect can be assessed by the SHAP values visualized by bar length and color (i.e. the larger the feature's share of the bar the more important; red represents rejecting effect and blue represents supporting effect).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

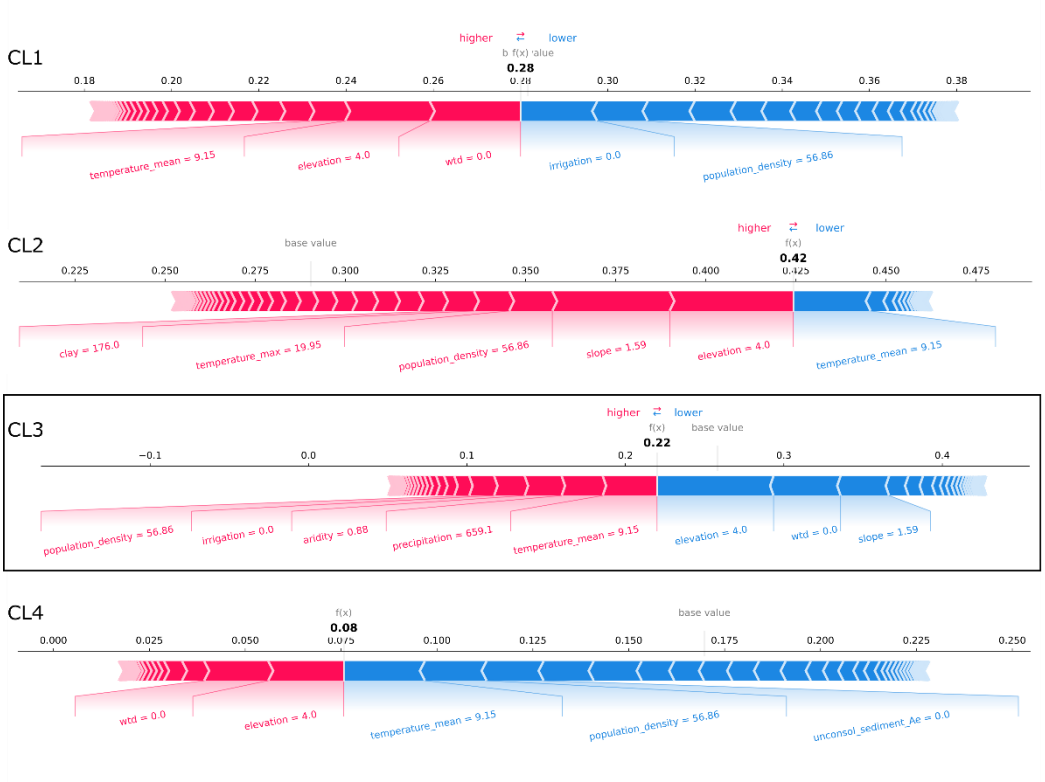


Figure E5 SHAP force plots explaining the prediction of the CL3 exemplary well site from the case study region in northern Germany (Fig. 5a) (from top to bottom: CL1, CL2, CL3, and CL4) that is confused with CL2. At this well site, the probability that CL2 is predicted is the largest in the case study (test dataset). The average scores of all classifications made by the Random Forest model for the training dataset are 0.28, 0.29, 0.26, and 0.17 respectively (base values). The model's probability scores for predicting the clusters at the exemplary well site (bold printed numbers) sum up to 100 %. Features that are important for the respective predictions are displayed with their values. Their importance and effect can be assessed by the SHAP values visualized by bar length and color (i.e. the larger the feature's share of the bar the more important; red represents rejecting effect and blue represents supporting effect).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

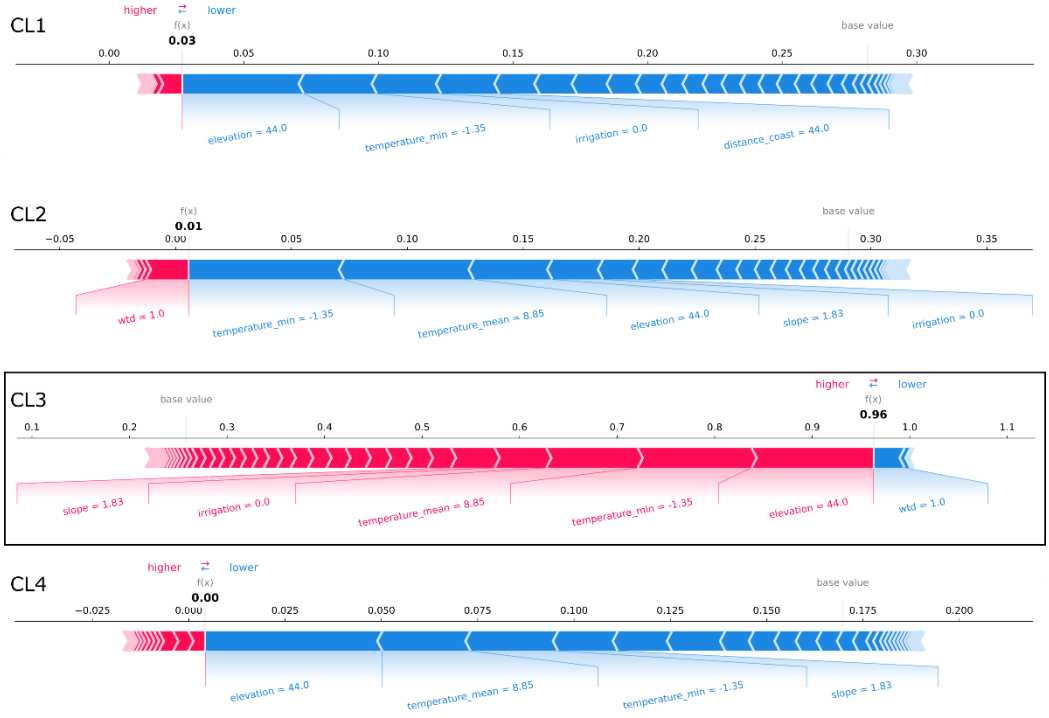


Figure E6 SHAP force plots explaining the correct prediction of the CL3 exemplary well site from the case study region in northern Germany (Fig. 5a) (from top to bottom: CL1, CL2, CL3, and CL4). At this well site, the probability that CL3 is predicted is the largest in the case study (test dataset). The average scores of all classifications made by the Random Forest model for the training dataset are 0.28, 0.29, 0.26, and 0.17 respectively (base values). The model's probability scores for predicting the clusters at the exemplary well site (bold printed numbers) sum up to 100 %. Features that are important for the respective predictions are displayed with their values. Their importance and effects can be assessed by the SHAP values visualized by bar length and color (i.e. the larger the feature's share of the bar the more important; red represents rejecting effect and blue represents supporting effect).

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

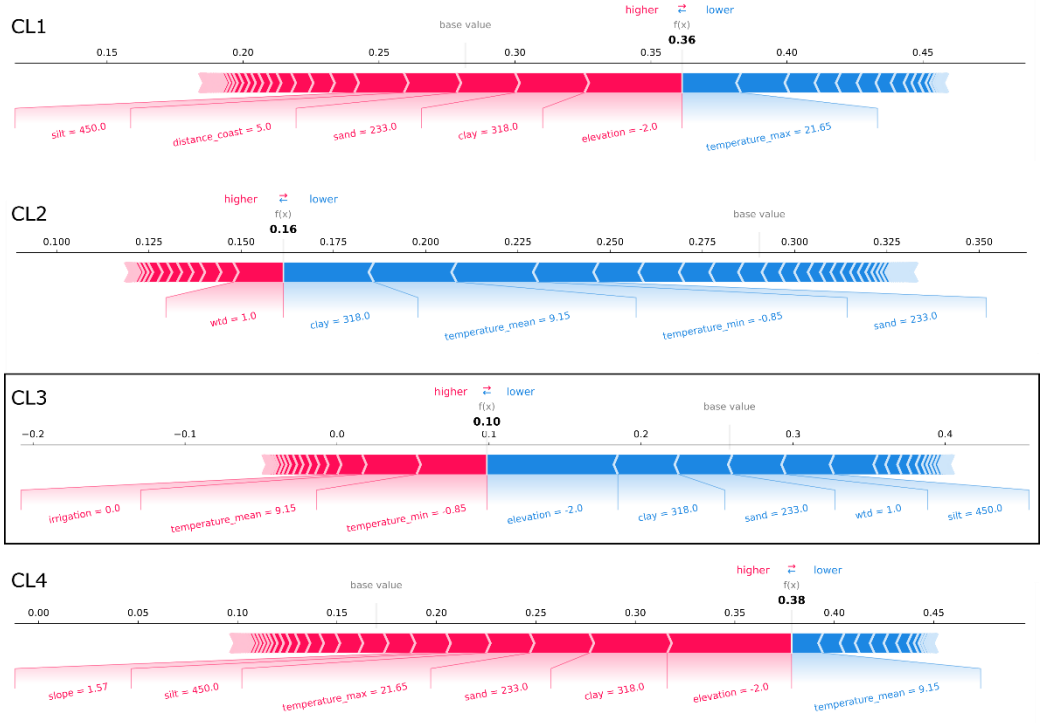


Figure E7 SHAP force plots explaining the prediction of the CL3 exemplary well site from the case study region in northern Germany (Fig. 5a) (from top to bottom: CL1, CL2, CL3, and CL4) that is confused with CL4. At this well site, the probability that CL4 is predicted is the largest in the case study (test dataset). The average scores of all classifications made by the Random Forest model for the training dataset are 0.28, 0.29, 0.26, and 0.17 respectively (base values). The model's probability scores for predicting the clusters at the exemplary well site (bold printed numbers) sum up to 100%. Features that are important for the respective predictions are displayed with their values. Their importance and effects can be assessed by the SHAP values visualized by bar length and color (i.e. the larger the feature's share of the bar the more important; red represents rejecting effect and blue represents supporting effect).

Code and data availability

We provide data to reproduce the results of this study (indices, attributes, clusters from k-means) at Zenodo (https://doi.org/10.5281/zenodo.8173404). Some of the data use agreements do not allow us to publish the original GWL time series and well locations. However, the groundwater data are available for free either via the web services or via request from governmental agencies listed in Table A1 (further information provided in the Supplement). Map data with information on attributes are available through the references listed in Table 1. The R code to calculate index values is available upon request from EH and BH. The Python code used for modelling and plotting is available upon request from AN. All the Python packages used in this research study (pandas, numpy, geopandas, sklearn, matplotlib, plotly, seaborn, scipy, shap) are freely available online.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Author contribution

AN collected and pre-processed the groundwater and attribute data, conceptualized the study, wrote the code, validated and visualized the results, and wrote the original paper draft. EH and BH contributed to the conceptualization. SB and JH supervised the work. All authors performed review and editing tasks.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a cooperation between the Climate Service Center Germany (GERICS) and Universität Hamburg, Germany, and conducted as part of the Future-H2O (Future climate and land-use change impacts on groundwater recharge rates and water quality for water resources) project.

Acknowledgments

The authors would like to thank all of the governmental agencies listed in Table A1 that supported this work by providing groundwater data and were available to answer questions about the data.

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, doi:10.5194/hess-21-5293-2017, 2017.

Ascott, M. J., Macdonald, D. M. J., Black, E., Verhoef, A., Nakohoun, P., Tirogo, J., Sandwidi, W. J. P., Bliefernicht, J., Sorensen, J. P. R., and Bossa, A. Y.: In Situ Observations and Lumped Parameter Model Reconstructions Reveal Intra-Annual to Multidecadal Variability in Groundwater Levels in Sub-Saharan Africa, Water Resour. Res., 56, D05109, doi:10.1029/2020WR028056, 2020.

Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H., and Schipper, A. M.: FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015, figshare [data set], doi:10.6084/m9.figshare.c.3890224.v1, 2018.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Barthel, R., Haaf, E., Giese, M., Nygren, M., Heudorfer, B., and Stahl, K.: Similarity-based approaches in hydrogeology: proposal of a new concept for data-scarce groundwater resource characterization and prediction, Hydrogeol. J., 24, 3392, doi:10.1007/s10040-021-02358-4, 2021.

Barthel, R., Haaf, E., Nygren, M., and Giese, M.: Systematic visual analysis of groundwater hydrographs: potential benefits and challenges, Hydrogeol. J., 30, 359–378, doi:10.1007/s10040-021-02433-w, 2022.

Baulon, L., Allier, D., Massei, N., Bessiere, H., Fournier, M., and Bault, V.: Influence of low-frequency variability on groundwater level trends, J. Hydrol., 606, 127436, doi:10.1016/j.jhydrol.2022.127436, 2022a.

Baulon, L., Massei, N., Allier, D., Fournier, M., and Bessiere, H.: Influence of low-frequency variability on high and low groundwater levels: example of aquifers in the Paris Basin, Hydrol. Earth Syst. Sci., 26, 2829–2854, doi:10.5194/hess-26-2829-2022, 2022b.

Beck, H. E., van Dijk, A. I. J. M., Roo, A. de, Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, doi:10.1002/2015WR018247, 2016.

Berendrecht, W., van Vliet, M., and Griffioen, J.: Combining statistical methods for detecting potential outliers in groundwater quality time series, Environ. Monit. Assess., 195, 85, doi:10.1007/s10661-022-10661-0, 2022.

Bergen, K. J., Johnson, P. A., Hoop, M. V. de, and Beroza, G. C.: Machine learning for data-driven discovery in solid Earth geoscience, Science, 363, doi:10.1126/science.aau0323, 2019.

Blumstock, M., Tetzlaff, D., Dick, J. J., Nuetzmann, G., and Soulsby, C.: Spatial organization of groundwater dynamics and streamflow response from different hydropedological units in a montane catchment, Hydrol. Process., 30, 3735–3753, doi:10.1002/hyp.10848, 2016.

Börker, J., Hartmann, J., Amann, T., and Romero-Mujalli, G.: Global Unconsolidated Sediments Map Database v1.0 (shapefile and gridded to 0.5° spatial resolution), PANGEA [data set], doi:10.1594/PANGAEA.884822, 2018.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001.

Caliński, T. and Harabasz, J.: A dendrite method for cluster analysis, Commun. Stat.-Theor. M., 3, 1–27, 1974.

Campello, R. J., Kröger, P., Sander, J., and Zimek, A.: Density-based clustering, Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 10, e1343, doi:10.1002/widm.1343, 2020.

CIESIN: Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11, NASA Socioeconomic Data and Applications Center (SEDAC) [data set], doi:10.7927/H49C6VHW, 2018.

Davies, D. L. and Bouldin, D. W.: A cluster separation measure, IEEE T. Pattern Anal., 224–227, doi:10.1109/TPAMI.1979.4766909, 1979.

Díaz-Alcaide, S. and Martínez-Santos, P.: Review: Advances in groundwater potential mapping, Hydrogeol. J., 27, 2307–2324, doi:10.1007/s10040-019-02001-3, 2019.

Donnelly, C., Ernst, K., and Arheimer, B.: A comparison of hydrological climate services at different scales by users and scientists, Clim. Serv., 11, 24–35, doi:10.1016/j.cliser.2018.06.002, 2018.

Elshall, A. S., Arik, A. D., El-Kadi, A. I., Pierce, S., Ye, M., Burnett, K. M., Wada, C. A., Bremer, L. L., and Chun, G.: Groundwater sustainability: a review of the interactions between science and policy, Environ. Res. Lett., 15, 93004, doi:10.1088/1748-9326/ab8e8c, 2020.

Esri: ArcGIS Pro, version 3.0.3, https://www.esri.com [software], 2022.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, Proc. Second Int. Conf. Knowl. Discov. Data Mining., 226–231, 1996.

Famiglietti, J. S.: The global groundwater crisis, Nat. Clim. Change, 4, 945–948, doi:10.1038/nclimate2425, 2014.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science, 339, 940–943, doi:10.1126/science.1229881, 2013.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., and Roth, L.: The shuttle radar topography mission, Rev. Geophys., 45, doi:10.1029/2005RG000183, 2007.

Ferguson, G. and Gleeson, T.: Vulnerability of coastal aquifers to groundwater use and climate change, Nat. Clim. Change, 2, 342–345, doi:10.1038/nclimate1413, 2012.

Giese, M., Haaf, E., Heudorfer, B., and Barthel, R.: Comparative hydrogeology – reference analysis of groundwater dynamics from neighbouring observation wells, Hydrolog. Sci. J., 65, 1685–1706, doi:10.1080/02626667.2020.1762888, 2020.

Gleeson, T., Cuthbert, M., Ferguson, G., and Perrone, D.: Global Groundwater Sustainability, Resources, and Systems in the Anthropocene, Annu. Rev. Earth Planet. Sci., 48, 431–463, doi:10.1146/annurev-earth-071719-055251, 2020.

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., Rahman, S., Oshinlaja, N., Maxwell, R., Lo, M.-H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J. S., Ducharne, A., Graaf, I. de, Cuthbert, M., Condon, L., Bresciani, E., and Bierkens, M. F. P.: GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545–7571, doi:10.5194/gmd-14-7545-2021, 2021.

Güler, C., Kurt, M. A., Alpaslan, M., and Akbulut, C.: Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques, J. Hydrol., 414-415, 435–451, doi:10.1016/j.jhydrol.2011.11.021, 2012.

Guppy, L., Uyttendaele, P., Villholth, K. G., and Smakhtin, V.: Groundwater and Sustainable Development Goals: Analysis of Interlinkages. UNU-INWEH Report Series, Issue 04. United Nations Unive, 26 pp., 2018.

Haaf, E., Giese, M., Heudorfer, B., Stahl, K., and Barthel, R.: Physiographic and Climatic Controls on Regional Groundwater Dynamics, Water Resour. Res., 56, doi:10.1029/2019WR026545, 2020.

Haaf, E., Giese, M., Reimann, T., and Barthel, R.: Data-driven Estimation of Groundwater Level Time-Series at Unmonitored Sites Using Comparative Regional Analysis, Water Resour. Res., doi:10.1029/2022WR033470, 2023.

Haaf, E. and Heudorfer, B.: Groundwater dynamics indices (0.1), Zenodo [data set], doi:10.5281/zenodo.1486058, 2018.

740 Haehnel, P., Rasmussen, T. C., and Rau, G. C.: Removing dynamic sea-level influences from groundwater-level measurements, Hydrol. Earth Syst. Sci. Discuss. [preprint], doi:10.5194/hess-2023-54, 2023.

Hartmann, J. and Moosdorf, N.: Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution), PANGAEA [data set], doi:10.1594/PANGAEA.788537, 2012.

Heudorfer, B., Haaf, E., Stahl, K., and Barthel, R.: Index-Based Characterization and Quantification of Groundwater
745 Dynamics, Water Resour. Res., 55, 5575–5592, doi:10.1029/2018WR024418, 2019.

Huggins, X., Gleeson, T., Serrano, D., Zipper, S., Jehn, F., Rohde, M. M., Abell, R., Vigerstol, K., and Hartmann, A.: Overlooked risks and opportunities in groundwatersheds of the world's protected areas, Nat. Sustain., 1–10, doi:10.1038/s41893-023-01086-9, 2023.

Huscroft, J., Gleeson, T., Hartmann, J., and Börker, J.: Compiling and mapping global permeability of the unconsolidated
750 and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). [Supporting Data], Borealis [data set], doi:10.5683/SP2/TTJNIU, 2018.

Johnson, T. D. and Belitz, K.: Assigning land use to supply wells for the statistical characterization of regional groundwater quality: correlating urban land use and VOC occurrence, J. Hydrol., 370, 100–108, doi:10.1016/j.jhydrol.2009.02.056, 2009.

755 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, Sci. Data, 4, 1–20, doi:10.1038/sdata.2017.122, 2017.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe, Earth Syst. Sci. Data, 13, 4529–4565, doi:10.5194/essd-13-4529-2021, 2021.

760 Knoll, L., Breuer, L., and Bach, M.: Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning, Sci. Total Environ., 668, 1317–1327, doi:10.1016/j.scitotenv.2019.03.045, 2019.

Knowling, M. J., Werner, A. D., and Herckenrath, D.: Quantifying climate and pumping contributions to aquifer depletion using a highly parameterised groundwater model: Uley South Basin (South Australia), J. Hydrol., 523, 515–530, doi:10.1016/j.jhydrol.2015.01.081, 2015.

765 Kulp, S. A. and Strauss, B. H.: CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network, Remote Sens. Environ., 206, 231–239, doi:10.1016/j.rse.2017.12.026, 2018.

Lall, U., Josset, L., and Russo, T.: A Snapshot of the World's Groundwater Challenges, Annu. Rev. Environ. Resour., 45, 171–194, doi:10.1146/annurev-environ-102017-025800, 2020.

Lee, S., Lee, K.-K., and Yoon, H.: Using artificial neural network models for groundwater level forecasting and assessment
770 of the relative impacts of influencing factors, Hydrogeol. J., 27, 567–579, doi:10.1007/s10040-018-1866-3, 2019.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process., 27, 2171–2186, doi:10.1002/hyp.9740, 2013.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Lehr, C. and Lischeid, G.: Efficient screening of groundwater head monitoring data for anthropogenic effects and measurement errors, Hydrol. Earth Syst. Sci., 501–513, doi:10.5194/hess-2019-287, 2020.

LfU-SH: Hydrogeologische Räume und Teilräume bezogen auf die oberflächennahen Wasserleiter, Landesamt für Umwelt des Landes Schleswig-Holstein, https://umweltportal.schleswig-holstein.de/trefferanzeige?docuuid=d7a7934c-d125-4fc4-93cc-08f34a2a3d7c, last access: 30 March 2023, 2003.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Sci. Data, 6, 283, doi:10.1038/s41597-019-0300-6, 2019.

Lischeid, G., Dannowski, R., Kaiser, K., Nützmann, G., Steidl, J., and Stüve, P.: Inconsistent hydrological trends do not necessarily imply spatially heterogeneous drivers, J. Hydrol., 596, 126096, doi:10.1016/j.jhydrol.2021.126096, 2021.

Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., and Hu, B. X.: Simulation of regional groundwater levels in arid regions using interpretable machine learning models, Sci. Total Environ., 831, 154902, doi:10.1016/j.scitotenv.2022.154902, 2022.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From Local Explanations to Global Understanding with Explainable AI for Trees, Nat. Mach. Intell., 2, 56–67, doi:10.1038/s42256-019-0138-9, 2020.

Mangor, K., Drønen, N. K., Kærgaard, K. H., and Kristensen, S. E.: Shoreline Management Guidelines, 4th ed., 462 pp., 2017.

Martens, K., van Camp, M., van Damme, D., and Walraevens, K.: Groundwater dynamics converted to a groundwater classification as a tool for nature development programs in the dunes, J. Hydrol., 499, 236–246, doi:10.1016/j.jhydrol.2013.06.045, 2013.

Martínez, M. L., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., and Landgrave, R.: The coasts of our world: Ecological, economic and social importance, Ecol. Econ., 63, 254–272, doi:10.1016/j.ecolecon.2006.10.022, 2007.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: A review, Hydrol. Process., 34, 1393–1409, doi:10.1002/hyp.13632, 2020.

Mishra, N., Khare, D., Gupta, K. K., and Shukla, R.: Impact of land use change on groundwater - a review, Adv. Water Resour. Prot., 2, 28–41, 2014.

Moeck, C., Grech-Cumbo, N., Podgorski, J., Bretzler, A., Gurdak, J. J., Berg, M., and Schirmer, M.: A global-scale dataset of direct natural groundwater recharge rates: A review of variables, processes and relationships, Sci. Total Environ., 717, 137042, doi:10.1016/j.scitotenv.2020.137042, 2020.

Moosdorf, N. and Oehler, T.: Societal use of fresh submarine groundwater discharge: An overlooked water resource, Earth-Sci. Rev., 171, 338–348, doi:10.1016/j.earscirev.2017.06.006, 2017.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Shadkam, S., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d - Standard model output, PANGAEA [data set], doi:10.1594/PANGAEA.918447, 2020.

Narvaez-Montoya, C., Mahlknecht, J., Torres-Martínez, J. A., Mora, A., and Bertrand, G.: Seawater intrusion pattern recognition supported by unsupervised learning: A systematic review and application, Sci. Total Environ., 864, 160933, doi:10.1016/j.scitotenv.2022.160933, 2023.

Nimmo, J. R., Perkins, K. S., Plampin, M. R., Walvoord, M. A., Ebel, B. A., and Mirus, B. B.: Rapid-Response Unsaturated Zone Hydrology: Small-Scale Data, Small-Scale Theory, Big Problems, Front. Earth Sci., 9, e12552, doi:10.3389/feart.2021.613564, 2021.

Nölscher, M., Mutz, M., and Broda, S.: Multiorder hydrologic Position for Europe - a Set of Features for Machine Learning and analysis in Hydrology, Sci. Data, 9, 662, doi:10.1038/s41597-022-01787-4, 2022.

Olden, J. D., Kennard, M. J., and Pusey, B. J.: A framework for hydrologic classification with a review of methodologies and applications in ecohydrology, Ecohydrol., 5, 503–518, doi:10.1002/eco.251, 2012.

Otto, R.: Estimating groundwater recharge rates in the southeastern Holstein region, northern Germany, Hydrogeol. J., 9, 498–511, doi:10.1007/s100400100155, 2001.

Oude Essink, G. H. P., van Baaren, E. S., and Louw, P. G. B. de: Effects of climate change on coastal groundwater systems: A modeling study in the Netherlands, Water Resour. Res., 46, doi:10.1029/2009WR008719, 2010.

Papacharalampous, G. and Tyralis, H.: Time Series Features for Supporting Hydrometeorological Explorations and Predictions in Ungauged Locations Using Large Datasets, Water, 14, 1657, doi:10.3390/w14101657, 2022.

Papacharalampous, G., Tyralis, H., Markonis, Y., and Hanel, M.: Hydroclimatic time series features at multiple time scales, J. Hydrol., 618, 129160, doi:10.1016/j.jhydrol.2023.129160, 2023.

Peng, J., Loew, A., Merlin, O., and Verhoest, N. E. C.: A review of spatial downscaling of satellite remotely sensed soil moisture, Rev. Geophys., 55, 341–366, doi:10.1002/2016RG000543, 2017.

Poggio, L., Sousa, L. M. de, Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, Soil, 7, 217–240, doi:10.5194/soil-7-217-2021, 2020.

Post, V., Kooi, H., and Simmons, C.: Using hydraulic head measurements in variable-density ground water flow analyses, Groundwater, 45, 664–671, doi:10.1111/j.1745-6584.2007.00339.x, 2007.

Probst, P., Wright, M. N., and Boulesteix, A.-L.: Hyperparameters and tuning strategies for random forest, WIREs Data Min. Knowl. Discov., 9, e1301, doi:10.1002/widm.1301, 2019.

Python Software Foundation: Python language reference, version 3.7.11 [software], https://www.python.org, 2021.

R Core Team: R: A language and environment for statistical computing [software], https://www.R-project.org/, 2021.

Rajaee, T., Ebrahimi, H., and Nourani, V.: A review of the artificial intelligence methods in groundwater level modeling, J. Hydrol., 572, 336–351, doi:10.1016/j.jhydrol.2018.12.037, 2019.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Rau, G. C., Cuthbert, M. O., Post, V. E. A., Schweizer, D., Acworth, R. I., Andersen, M. S., Blum, P., Carrara, E., Rasmussen, T. C., and Ge, S.: Future-proofing hydrogeology by revising groundwater monitoring practice, Hydrogeol J, 28, 2963–2969, doi:10.1007/s10040-020-02242-7, 2020.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, doi:10.1038/s41586-019-0912-1, 2019.

Retike, I., Bikše, J., Kalvāns, A., Dēliņa, A., Avotniece, Z., Zaadnoordijk, W. J., Jemeljanova, M., Popovs, K., Babre, A., and Zelenkevičs, A.: Rescue of groundwater level time series: How to visually identify and treat errors, J. Hydrol., 605, 127294, doi:10.1016/j.jhydrol.2021.127294, 2022.

Riedel, T. and Weber, T. K. D.: Review: The influence of global change on Europe's water cycle and groundwater recharge, Hydrogeol. J., 28, 1939–1959, doi:10.1007/s10040-020-02165-3, 2020.

Rinderer, M., Meerveld, H. J., and McGlynn, B. L.: From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics, Water Resour. Res., 55, 5784–5806, doi:10.1029/2018WR023886, 2019.

Rodriguez, E., Morris, C. S., and Belz, J. E.: A global assessment of the SRTM performance, Photogramm. Eng. Remote Sens., 72, 249–260, doi:10.14358/PERS.72.3.249, 2006.

Rousseeuw, P. J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, doi:10.1016/0377-0427(87)90125-7, 1987.

Sayre, R., Karagulle, D., Frye, C., Boucher, T., Wolff, N. H., Breyer, S., Wright, D., Martin, M., Butler, K., and van Graafeiland, K.: An assessment of the representation of ecosystems in global protected areas using new maps of World Climate Regions and World Ecosystems, Glob. Ecol. Conserv., 21, doi:10.1016/j.gecco.2019.e00860, 2020.

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, J. Adv. Model. Earth Syst., 9, 65–88, doi:10.1002/2016MS000686, 2017.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., and Tsai, W.-P.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, Hydrol. Earth Syst. Sci., 22, 5639–5656, doi:10.5194/hess-22-5639-2018, 2018.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: Historical Irrigation Dataset (HID), MyGeoHUB [data set], doi:10.13019/M20599, 2015.

Smerdon, B. D.: A synopsis of climate change effects on groundwater recharge, J. Hydrol., 555, 125–128, doi:10.1016/j.jhydrol.2017.09.047, 2017.

Sorensen, J. P. R., Davies, J., Ebrahim, G. Y., Lindle, J., Marchant, B. P., Ascott, M. J., Bloomfield, J. P., Cuthbert, M. O., Holland, M., Jensen, K. H., Shamsudduha, M., Villholth, K. G., MacDonald, A. M., and Taylor, R. G.: The influence of groundwater abstraction on interpreting climate controls and extreme recharge events from well hydrographs in semi-arid South Africa, Hydrogeol. J., 29, 2773–2787, doi:10.1007/s10040-021-02391-3, 2021.

https://doi.org/10.5194/hess-2023-180
Preprint. Discussion started: 24 August 2023

Trabucco, A. and Zomer, R.: Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2, figshare [data set], doi:10.6084/m9.figshare.7504448.v3, 2019.

United Nations: The United Nations World Water Development Report 2022: Groundwater - Making the invisible visible, doi:10.18356/9789210015363, 2022.

Vahdat-Aboueshagh, H., Tsai, F. T.-C., Bhatta, D., and Paudel, K. P.: Irrigation-Intensive Groundwater Modeling of Complex Aquifer Systems Through Integration of Big Geological Data, Front. Water, 3, doi:10.3389/frwa.2021.623476, 2021.

Wang, X., Smith, K., and Hyndman, R.: Characteristic-Based Clustering for Time Series Data, Data Min. Knowl. Disc., 13, 335–364, doi:10.1007/s10618-005-0039-x, 2006.

Winkler, K., Fuchs, R., Rounsevell, M. D. A., and Herold, M.: HILDA+ Global Land Use Change between 1960 and 2019, PANGAEA [data set], doi:10.1594/PANGAEA.921846, 2020.

Worland, S. C., Steinschneider, S., Asquith, W., Knight, R., and Wieczorek, M.: Prediction and Inference of Flow Duration Curves Using Multioutput Neural Networks, Water Resour. Res., 55, 6850–6868, doi:10.1029/2018WR024463, 2019.

Wriedt, G.: Verfahren zur Analyse klimatischer und anthropogener Einflüsse auf die Grundwasserstandsentwicklung, Grundwasser, 22, 41–53, doi:10.1007/s00767-016-0349-5, 2017.

Wunsch, A., Liesch, T., and Broda, S.: Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles, Water Resour. Manag., doi:10.1007/s11269-021-03006-y, 2021.

Wunsch, A., Liesch, T., and Broda, S.: Deep learning shows declining groundwater levels in Germany until 2100 due to climate change, Nat. Commun., 1–13, doi:10.1038/s41467-022-28770-2, 2022.

Yang, Y. and Chui, T. F. M.: Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods, Hydrol. Earth Syst. Sci., 25, 5839–5858, doi:10.5194/hess-25-5839-2021, 2021.