



Feature-based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing Map-Ensembles

Andreas Wunsch¹ · Tanja Liesch¹ · Stefan Broda²

Received: 26 October 2020 / Accepted: 16 October 2021 / Published online: 28 October 2021
© The Author(s) 2021

Abstract

Hydrograph clustering helps to identify dynamic patterns within aquifers systems, an important foundation of characterizing groundwater systems and their influences, which is necessary to effectively manage groundwater resources. We develop an unsupervised modeling approach to characterize and cluster hydrographs on regional scale according to their dynamics. We apply feature-based clustering to improve the exploitation of heterogeneous datasets, explore the usefulness of existing features and propose new features specifically useful to describe groundwater hydrographs. The clustering itself is based on a powerful combination of Self-Organizing Maps with a modified DS2L-Algorithm, which automatically derives the cluster number but also allows to influence the level of detail of the clustering. We further develop a framework that combines these methods with ensemble modeling, internal cluster validation indices, resampling and consensus voting to finally obtain a robust clustering result and remove arbitrariness from the feature selection process. Further we propose a measure to sort hydrographs within clusters, useful for both interpretability and visualization. We test the framework with weekly data from the Upper Rhine Graben System, using more than 1800 hydrographs from a period of 30 years (1986–2016). The results show that our approach is adaptively capable of identifying homogeneous groups of hydrograph dynamics. The resulting clusters show both spatially known and unknown patterns, some of which correspond clearly to external controlling factors, such as intensive groundwater management in the northern part of the test area. This framework is easily transferable to other regions and, by adapting the describing features, also to other time series-clustering applications.

Keywords Groundwater-Dynamics · Time series clustering · Machine learning · Self-organizing maps (SOM) · Ensemble modeling · Feature clustering

1 Introduction

The analysis and evaluation of groundwater level dynamics can contribute valuable information to assess quantitative groundwater availability, which is important to manage groundwater resources and secure water supply in many regions worldwide. As every

✉ Andreas Wunsch
andreas.wunsch@kit.edu

Extended author information available on the last page of the article

hydrograph contains information about system properties (e.g. geology), artificial (e.g. withdrawal) and natural (e.g. streamflow interaction) environmental factors, hydrograph clustering is often helpful to identify common dynamics and to differentiate between signals resulting from external controlling factors and noise. This improves understanding of system dynamics, and forms the basis for further analysis including forecasting or scenario building. Popular methods for clustering hydrological time series are for example Cluster-Analysis (CA) (Naranjo-Fernández et al. 2020) and PCA (Haaf and Barthel 2018), each alone or as a combination of both (Machiwal and Singh 2015). Besides classical approaches, Artificial Neural Networks (ANN) offer innovative concepts to deal with larger sets of multidimensional data, for example by using Self-Organizing Maps (SOM) for unsupervised clustering. Several studies from different disciplines compare SOM to other well-established clustering methods like k-means and hierarchical clustering (HC). Some authors found that k-means performs equally (He et al. 2004) or even better than SOM (Balakrishnan et al. 1994; Kumar and Dhamija 2010; Mingoti and Lima 2006); however, there is no consent on this aspect in the literature as other authors found SOM to be clearly superior to k-means (Chen et al. 2010; Kiang et al. 2006; Melo Riveros et al. 2019) and also to HC (Mangiameli et al. 1996). Often, SOM are even combined with k-means or HC methods, because interpreting a trained SOM structure is not trivial and usually second-level clustering is therefore applied. Besides classical clustering methods, also algorithms specialized on the interpretation of trained SOM, such as DS2L (Cabanes et al. 2012), exist. In the hydrological context, SOM have been extensively used to analyze water quality and chemistry (Gholami et al. 2021). Applications to groundwater hydrographs are forecasting by using hybrid SOM-ANN models (Chang et al. 2016; Chang et al. 2014; Chen et al. 2010; Lin and Chen 2005; Moradkhani et al. 2004), hydrological event type clustering and classification (Abrahart and See 2000; Toth 2009), or catchment classification (Toth 2013). The clustering of groundwater hydrographs, especially by using SOM, has been carried out rather rarely so far. Han et al. (2016) used SOM to identify homogeneous clusters of groundwater level piezometers as a preprocessing step to forecasting with a step-wise cluster multi-site inference model. However, they tested the approach on a rather small number of wells (30) and more importantly, they used the time series directly as inputs. Approaches that use time series directly for clustering suffer from dependency on high-quality data (equal length, equal period, no gaps). Application of feature-based approaches can overcome this problem by using patchy input data (Wang et al. 2006). Features, in this case, are descriptive (statistical) measures of the time series, extracted e.g. from the time or frequency domain (Caiado et al. 2015). To apply a feature-based approach on groundwater level data, features taking the peculiarities of groundwater hydrographs into account are desirable. Heudorfer et al. (2019) present a comprehensive compilation of 45 possibly suited indices to describe groundwater dynamics. Their approach is very much related to the concept of hydrological signatures (McMillan et al. 2017), where features are designed to describe certain dynamic aspects in surface hydrology. Feature-based clustering of hydrological time series using Self-Organizing Maps has already been performed by Nourani et al. (2015), who used features based on wavelet decomposition to cluster a small number of wells on Ardabil plain, Iran. However, to the best knowledge of the authors, no approach is known yet that combines SOM-clustering with specifically designed features that describe the dynamic aspects of certain groundwater hydrographs.

In this study, we develop a robust, flexible, and semi automated framework for groundwater hydrograph clustering. We chose feature-based time series clustering, which allows

to use data from time series of different periods, different lengths as well as missing and noisy data. Moreover, we present and explore several new features, which showed promising results and which are particularly suited to describe the dynamic aspects of groundwater hydrographs. We introduce a modification of a powerful clustering algorithm combination (SOM+DS2L) that allows influence on the level of detail of the clustering result, and implement Ensemble-Modeling-Techniques to remove arbitrariness from the feature selection process as well as to ensure a higher robustness of the clustering result. We apply the developed approach to the Upper Rhine Graben (URG) area in central Europe, based on a dataset of overall 1853 groundwater hydrographs. The motivation and later application is the reduction of the forecasting workload of regional forecasting of groundwater levels by selecting representative hydrographs from the clustering result. Additionally, we aim for increased system understanding in terms of dynamic patterns and their main controlling factors.

2 Data and Study Area

The study area is the Upper Rhine Graben (URG), mainly located in southwestern Germany and northeastern France (Fig. 1a). It is the largest groundwater resource in central Europe (LUBW 2006), covering 80% of the drinking water demand of the region (Région Alsace - Strasbourg 1999) and is also intensively used for water extraction for irrigation and industrial purposes. The URG, a Cenozoic rift structure, 300 km long (N-S) and on average 40 km wide (E-W), is filled with sediments (mainly gravel and sand) with a total thickness of up to about 3500 m. Hydrogeologically, the uppermost Quaternary sediments are most important. They reach a thickness of more than 200 m in the southern part, which strongly decreases to about 30 m in the area around Karlsruhe. In the northern part of the URG, the Quaternary sediment thickness increases to up to 500 m and a multi-aquifer system exists due to several fine-clastic layers dividing the Quaternary sediments (Geyer et al. 2011; LUBW 2006).

The dataset used consists of 1853 weekly groundwater hydrographs from Germany and France, including one synthetic hydrograph with strong outlier characteristics to explore and illustrate additional properties of the clustering approach. The considered period ranges from October 1986 to September 2016 (30 years). The majority of the hydrographs show data for almost the entire period, the shortest length included being six years. We removed strong outliers conservatively and interpolated small data gaps to up to one month linearly. Figure 1a shows the study area in general (left) and the locations of the 1852 real wells included in the dataset (right). The dataset includes only wells from the uppermost aquifer within the Quaternary sediments, which causes e.g. the three major blank spaces on the map in Fig. 1a (right) due to locally changing geological conditions in these areas.

Figure 1b sketches a strongly simplified E-W cross-section of the URG and illustrates that the regional groundwater dynamics are the result of a complex interaction of multiple factors, which we divided into processes (Pr), driving forces (DF), and governing parameters (GP) for the sake of a more systematic point of view. Processes are the physical processes that directly influence the groundwater levels (e.g. recharge). They are mostly driven by external driving forces (e.g. precipitation) and in most cases depend on one or several governing parameters (e.g. topography, land use). A detailed assessment of the importance of each factor can be found in the electronic supplement (Text S1).

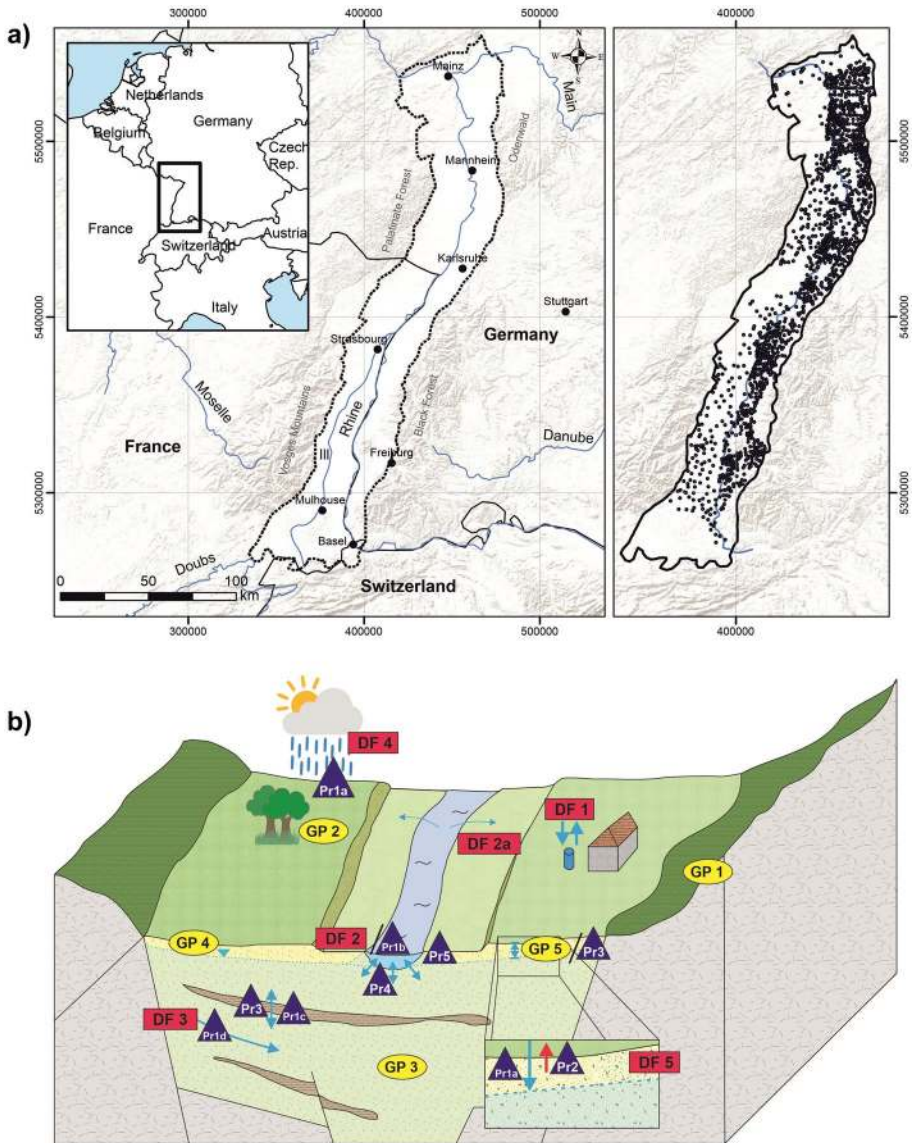


Fig. 1 (a) The study area Upper Rhine Graben (left) and the locations of the used 1852 real groundwater monitoring wells (right). (b) Strongly simplified E-W cross-section of the URG, summarizing some influences on groundwater dynamic patterns (DF: driving force, GP: governing parameter, Pr: process); DF1 – artificial extraction/infiltration, DF2 – surface water interactions (a: floods), DF3 – regional flow systems, DF4 – weather/climate, DF5 – soil moisture; GP1 – topography, GP2 – vegetation/land use, GP3 – geology (aquifer type/material properties), GP4 – pressure state (free/confined), GP5 – mean depth to groundwater; Pr1 – recharge (a: direct/diffuse; b: direct/local; c: inter-aquifer-exchange; d: lateral), Pr2 – evapotranspiration, Pr3 – signal damping (low pass filter effect), Pr4 – in-/exfiltration, Pr5 – bank storage

3 Methodology

3.1 Feature-Based Time Series Characterization

A proper feature set, depending on the unique hydrogeological conditions, is key to adequately describe and thus successfully cluster the data. Here, features are descriptive (statistical) indices that quantify the dynamics of groundwater hydrographs, similar to the concept of signatures in hydrology (McMillan et al. 2017). However, groundwater hydrographs generally differ considerably from surface water hydrographs, which makes many hydrological signatures inadequate for describing dynamic aspects of groundwater. Thus, there is a need for comprehensive testing of the transferability to the groundwater domain, as was done by Heudorfer et al. (2019). A most important supportive tool for pre-selecting adequate features is a visual skill test to check the adequacy and the explanatory power of every single feature. Applying PCA or related methods can help to reduce the feature number by ruling out redundant features based on the explained variance. However, including correlated features can help to improve the result, by up-weighting important aspects of the general dynamics. We explore this aspect with a correlation analysis of all selected features in the results section. In total, we tested a broad variety of feature candidates (> 50), including standard statistics measures, features derived from the literature (Heudorfer et al. 2019; Wang et al. 2006), as well as self-designed features to account for peculiarities of both the study area and groundwater hydrographs in general. In the following, we introduce those which have successfully passed the visual skill test for our data set. Skill test results that show the explanatory power of each feature are provided in the supplementary material (Figs. S1 to S13). Table 1 summarizes the feature calculation, the corresponding data basis, and the primary purpose or a short description for all used features. For more details on the self-designed features, we refer to the supplementary material where we also present results on the robustness of the features against gaps, noise and time series length (Text S2 and S4, Tables S1–S3, Figs. S1–S13).

3.2 Self-Organizing Map Clustering Using DS2L Algorithm

SOM perform a non-linear projection of multidimensional data onto a regular neuron lattice surface. They show characteristics of both clustering (local averaging) and data compression methods (topology preservation), which is a unique property and also an advantage of SOM compared to other cluster algorithms and projection methods (Kohonen 2014). Every neuron has clearly identifiable neighbors, which allows simple two-dimensional visual representations of multi-dimensional data. We apply a modified version of the density-based simultaneous two-level (DS2L)-algorithm (Cabanes et al. 2012) to automatically derive clusters from the trained SOM. DS2L detects clusters by analyzing data density and neighborhood connection-strength of the SOM. An adequate cluster number is automatically determined and the algorithm does not tend to produce clusters of equal size, both advantages compared to some well-established cluster algorithms (e.g. k-means or some hierarchical methods). We modify DS2L-algorithm in such a way that the user can decide purely qualitatively whether the clustering should be performed more coarsely or more finely. On the chosen level of detail the cluster number is still determined automatically. For this, we implement three adjustment parameters for thresholds of data density and neighborhood connection-strength as well as to control the application of some algorithm steps. Besides

Table 1 List of promising features (passed skill test) to describe the groundwater dynamics of time series in the URG dataset. Features in italics were not used based on the decision of the ensemble approach (see Sects. 3.1/3.3)

Feature Name (Abbrev.)	Data*	Purpose / Description	Ref**
Range Ratio (RR)	o	Detection of superimposed long-periodic signals, also sensitive to outliers, calculated as the ratio of the mean annual range to the overall range	sd
Skewness (Skew)	o	Boundedness, inhomogeneities, outliers, asymmetry of the probability distribution	ss
Annual Periodicity (P52)	o	Strength of the annual cycle, calculated by correlating (Pearson) the mean annual (52 weeks) periodicity with the complete time series	sd
SDdiff	o	Flashiness, frequency, and rapidity of short-term changes, calculated as the standard deviation of all first derivatives	sd
Longest Recession (LRec)	o	(unnaturally) long descending heads, longest sequence without rising head values	sd
Jumps	z	Inhomogeneities/breaks, partly also variability, calculated as the absolute and standardized maximum change of the mean of two successive years	sd
Seasonal Behaviour (SB)	z	Position of the maximum in the annual cycle, agreement with the expected average seasonality (Min in September, Max in March)	sd
Median[0,1] (Med01)	n	Boundedness, median after scaling to [0,1], standard statistics measure, derived from (Heudorfer et al. 2019)	ss/lit
High Pulse Duration (HPD)	n	Average duration of heads exceeding the 80th percentile of non-exceedance, for details see Richter et al. (1996), derived from (Heudorfer et al. 2019)	lit
<i>Richards-Baker Index (RBI)</i>	<i>o</i>	<i>Flashiness, frequency, and rapidity of short-term changes, for detailed explanation see Baker et al. (2004)</i>	<i>lit</i>
<i>Yearly Variance (Yvar)</i>	<i>z</i>	<i>Variability, periodicity, calculated as the median of the yearly calculated variances</i>	<i>sd</i>
<i>Standard Error of the Mean (SEM)</i>	<i>o</i>	<i>Standardized statistical dispersion, calculated as the standardized standard deviation of the time series</i>	<i>ss</i>
<i>Low Pulse Duration (LPD)</i>	<i>n</i>	<i>Average duration of heads dropping below the 20th percentile of non-exceedance, for details see Richter et al. (1996), derived from Heudorfer et al. (2019)</i>	<i>lit</i>

* o: original, z: z-scored, n: normalized

** lit: literature, sd: self-designed, ss: standard statistics

the number of neurons (SOM-size), which also has an influence on the cluster result, the following four parameters must be optimized during the clustering process.

- **SOM-size:** *normal* ($5\sqrt{n}$), *small* ($5\sqrt{n} \cdot 0.25$) or *big* ($5\sqrt{n} \cdot 4$) - options implemented in SOM-Toolbox (Vesanto 2005), n: number of samples
- **NTH:** $NTH \geq -1 \in \mathbb{Z}$ - DS2L-Neighborhood-Threshold, connection strength required to qualify as cluster border, -1 means connection strength is not used.
- **DR:** Yes/No - DS2L-Density-Refinement, use density values for cluster determination
- **DM:** Yes/No - DS2L-Density-Merging, merge similar clusters based on density-dependent index

3.3 Workflow

Figure 2 summarizes the workflow of the approach applied in this study. A common problem with many feature-based approaches is the arbitrariness of feature selection. As shown by line I in Fig. 2, we implement an SOM-ensemble to find the best combination of all pre-selected features, whereby the cluster quality is judged by five different internal validation indices (Caliński-Harabasz criterion (CH), McClain-Rao criterion (MR), PBM-Index, Ratkowsky-Lance criterion (RL), C-Index). Line II in Fig. 2 shows a second SOM-Ensemble based on delete-d-jackknifing resampling. Its purpose is to simulate changes in the observational network by manipulating the input data set, and to obtain cluster results as robust as possible. The final cluster result is based on voting consensus. For visualization and evaluation, we rearrange all original time series of a cluster by their mean pairwise Pearson-correlation with all other cluster members. A weighting by the p-value of the respective single correlations lowers correlation values with low significance (which might arise from only short overlapping time periods). We define this value as the weighted intra-cluster correlation (\overline{R}_w). A detailed description and discussion of the workflow is added to the supplementary material (Text S3).

Besides the clustering itself, interpreting the results is very useful to improve system understanding in general. This is especially the case for clusters, which are not easily interpretable in terms of spatial location or dynamic aspects. Hence, we conduct detailed correlation analyses for factors mentioned in Fig. 1b, where reasonable additional data are available to perform meaningful statistics. For some, data are only available for part of the study area. We therefore link them also with features and not only with clusters. In this way, we avoid a bias, for clusters with wells in areas without data. Furthermore, the dynamics within clusters are usually the result of a superposition of several influencing factor which

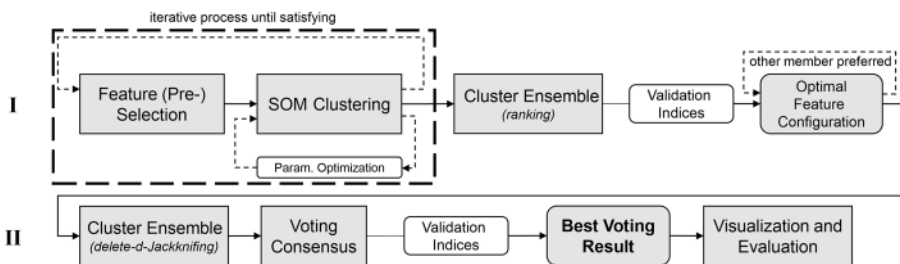


Fig. 2 Workflow of the presented methodology

can make correlations rather challenging. Because of the easier metric interpretation, we focus on linear correlation analysis, although we are aware that non-linear relationships can also exist. In addition, we only mention significant correlations with $p \leq 0.05$.

4 Results and Discussion

We applied our approach to 1853 time series from the Upper Rhine Graben area (including one synthetic hydrograph). The feature pre-selection provided 13 features with good explanatory power regarding our specific dataset (Sect. 3.1/Table 1). The used cluster parameter combination was: SOM-size: big, NTH = 0, DR: Yes, DM: No (Sect. 3.2). The best feature configuration derived from the first ensemble (115.005 members) included 9 out of 13 features.

As stated in Sect. 3.1, we found that including correlated features improves the clustering results. A correlation analysis among the included features shows the highest absolute significant ($p < 0.05$) correlations for the features Skew-Med01 (-0.81) and P52-RR (0.79), which is consistent with the meaning and calculation of these respective feature pairs (e.g. hydrographs with high annual periodicity often also show a regular range over the years, thus high RR values). A detailed correlation matrix of all features can be found in the supplementary material (Fig. S27).

The final cluster result consists of 18 clusters (Fig. 3a) with sizes ranging from 239 hydrographs in cluster 1, to only one hydrograph in cluster 18, which is the synthetic hydrograph with outlier characteristics (cluster numbers sorted in descending order by size). The five biggest clusters include almost 1000 of the 1853 hydrographs in total, eight clusters show sizes larger than 100, only five clusters show sizes below 50. Due to the huge amount of information, we summarize detailed information and graphics on every single cluster in the supplement (Figs. S28–S65). In the following, we only present selected results.

The Boxplots in Fig. 3b show the feature value distributions within each cluster. For some clusters a clear feature importance can be derived. Cluster 2, for example, is comprised of mainly regular hydrographs dominated by the annual periodicity and with little other long- or short-term periodicities (high P52), as well as the annual maximum and minimum occurring very regularly during March and September, respectively (high SB). Reasons are comparably high recharge values in the middle of the Graben, typical for wells neither strongly dominated by margin inflows nor by the Rhine River. However, less straightforward feature combinations also exist which are therefore harder to interpret. The same applies to the spatial distribution of the clusters. If there is no distinct grouping (e.g. as a result of a spatially limited, local influence on the dynamics), more effort is required to understand what processes, forces, or parameters might be the cause of the common dynamics.

Cluster 3 (Fig. 4) is an example of straightforward interpretation, where wells follow almost exclusively the Rhine River course. Thus, identifying interaction with surface water (DF2, Pr1b, Pr4, P5, Fig. 1b) as the dominant driving force is comparatively easy. Some wells of this cluster showing greater distances to the Rhine River are in turn closer to mid-sized rivers like the Neckar or Ill, where common dynamics can be expected due to similar overall conditions. The resulting hydrographs grouping reveals that despite data gaps and different time series lengths, still a homogeneous grouping was achieved by our approach. The weighted intra-cluster correlation values (R_w) are expressed by the coloring (the brighter the lower), thus by the sorting of the stacked time series and by the bars on the

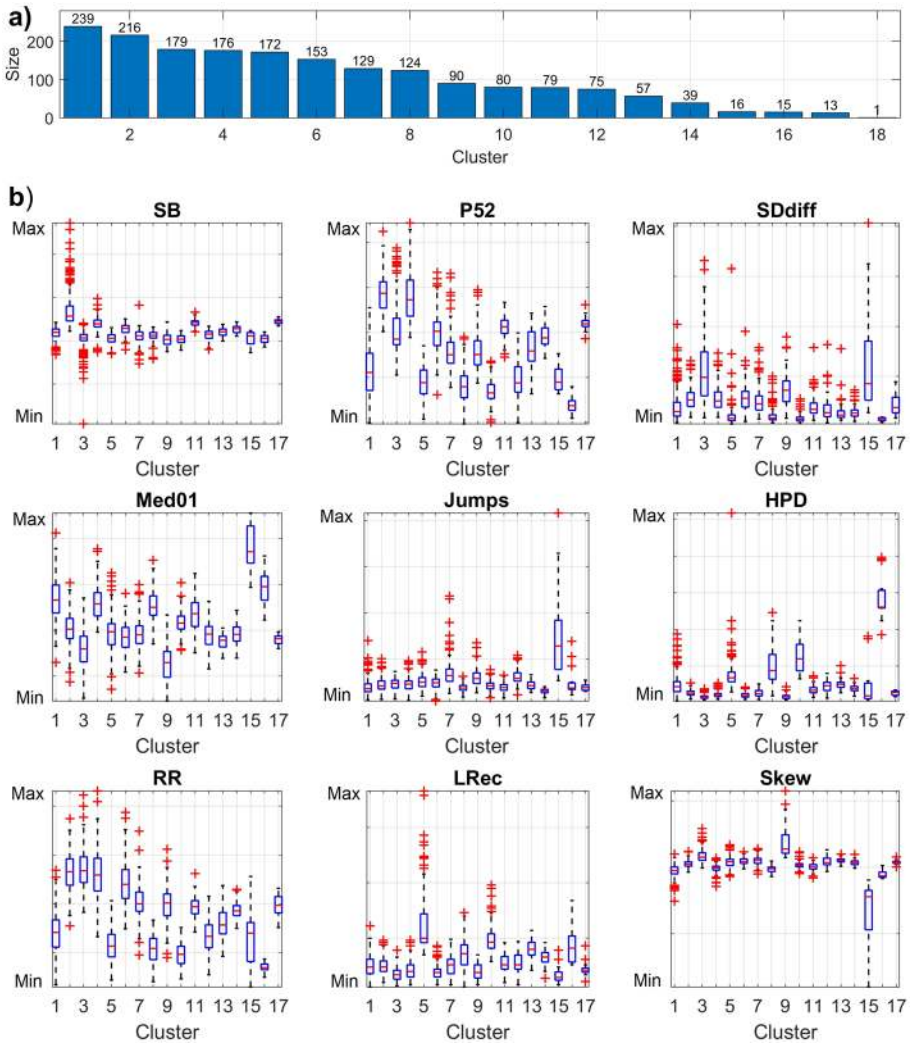


Fig. 3 a) cluster sizes, b) feature value boxplots of all clusters. For a better graphical representation, Cluster 18 was omitted due to strong outlier characteristics. Boxplots including Cluster 18 can be found in the supplement

right. In general, with decreasing (\overline{R}_w)-values towards the cluster borders the heterogeneity increases and the certainty of the cluster assignment of individual hydrographs decreases. Considering cluster 3, we can observe a distinct north-south gradient, which means that despite an changing dynamic along the river, grouping was still successful. However, other wells close to the Rhine River were sorted into different clusters, but show indeed different dynamics (compare clusters 7 and 9 in the supplement). In terms of feature values, the Rhine influence for cluster 3 is best expressed by feature SDdiff, describing the higher flashiness close to the river (Fig. 3). Other features are also in accordance. For example, Med01 values are comparably low, indicating that the hydrographs are more likely to be

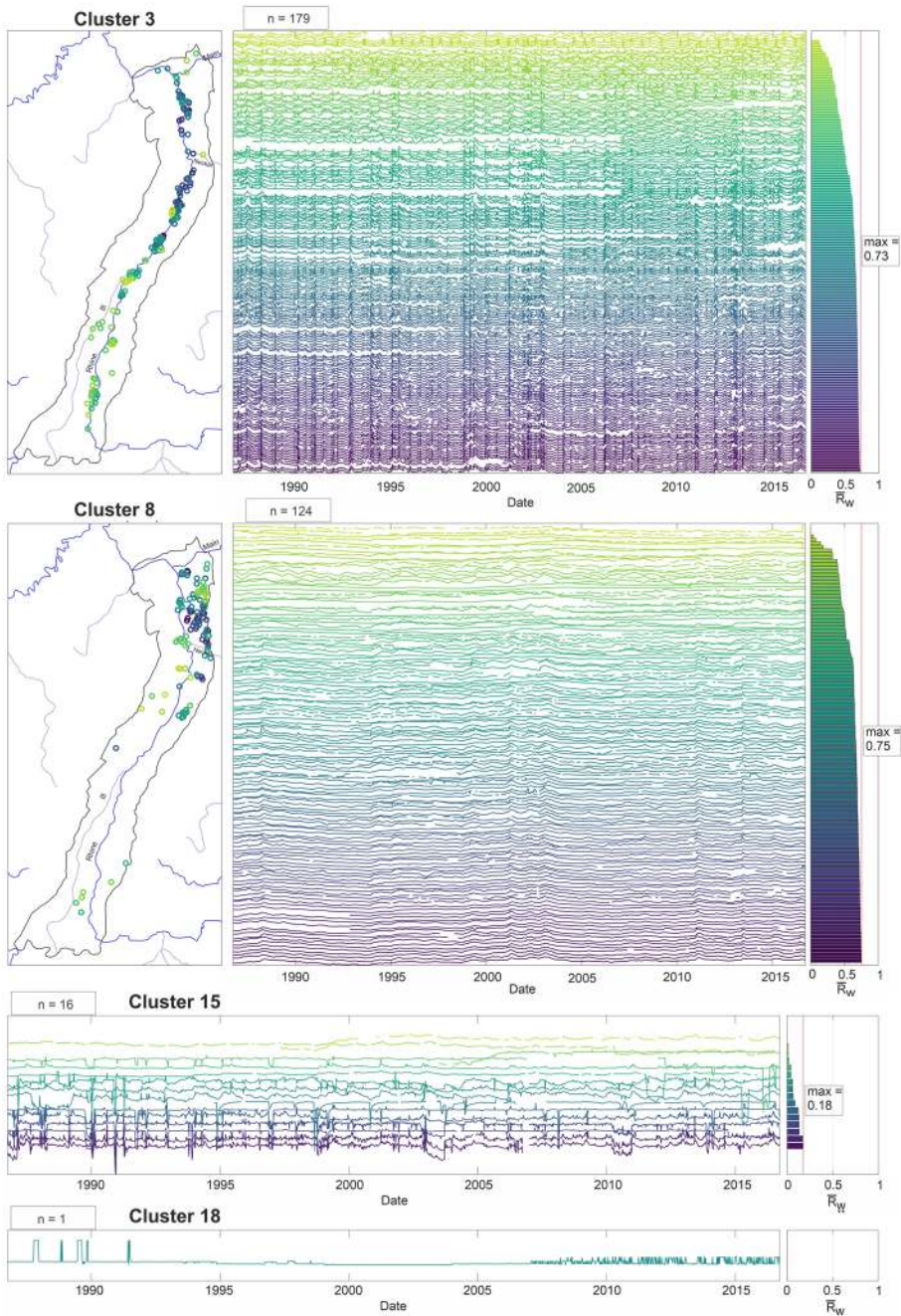


Fig. 4 Maps and stacked z-transformed hydrographs of selected clusters. Coloring and stacked order reflect the weighted intra-cluster correlation (R_W) also shown as bar-plot on the far right; Cluster 3 is mainly influenced by the Rhine River; Cluster 8 shows spatial grouping in the northern part and contains hydrographs with low annual periodicity and low variability; Cluster 15 groups hydrographs with outliers and inhomogeneities; Cluster 18 contains only the synthetic hydrograph, which is a heavy outlier compared to the whole dataset

bound to some kind of baseflow level in combination with short and high peaks triggered by the streamflow.

Overall results show that in the north of the URG, predominantly hydrographs with small variability and weak annual periodicity occur, while especially the middle section of the URG shows highly seasonal and highly regular hydrograph patterns. The former is expressed mainly by clusters 1, 5, 8, 10, 16; the latter can be seen e.g. in clusters 2 and 4 (Figs. S29–S65). We selected cluster 8 (Fig. 4) to illustrate the low-variance case in the northern URG. Driving forces connected to this cluster are most certainly strong anthropogenic influences (DF1, Fig. 1b) because the cluster focuses spatially on an area with strong groundwater management efforts. Connections to in generally lower groundwater recharge values (Pr1a, Fig. 1b) in the northern URG can also be drawn. Both factors can explain the smoothness as well as the comparatively weak annual periodicity and low variability of the hydrographs in cluster 8.

The approach successfully separates a small group of 16 hydrographs with outliers and significant inhomogeneities, which probably occur due to two major Rhine River weir locks (Strasbourg, Breisach) (cluster 15, Fig. 4). Furthermore, the synthetic hydrograph is put in a separate cluster (cluster 18, Fig. 4). Such clusters are rather based on single events or characteristics than on similar, highly correlated time series. Therefore, even for good clusters in terms of such events, (R_W) -values can be rather low.

In terms of system understanding, thus the correlation analysis of clusters and features with explaining factors, we found that the mean depth to groundwater (GP5, Fig. 1b) shows clear negative correlations (P52 (-0.45), RR (-0.44), SB (-0.29), SDdiff (-0.16)) with features describing the variability of hydrographs (e.g. seasonality, flashiness). Such variability is generally damped with increasing depth to groundwater. The complementary case applies to HPD (0.33) and LRec (0.29), which both rather reach higher values for smoother hydrographs with little short-term variations. A clear relation to the clusters could not be found, though, probably due to the only minor variation of this parameter (70% of the wells < 5 m bgl on average), which makes a meaningful interpretation of the cluster development challenging. We observed only slight tendencies to greater or smaller depths to GW for some clusters. Another probable explanation could be that more dominating factors superimpose the effect of the depth to groundwater and are thus more decisive for cluster assignment.

We explored the connection of features and clusters to diffuse groundwater recharge (Pr1a, Fig. 1b) using GWN1000 data (BGR 2019). French wells (190) were excluded due to no data. In accordance with the findings and explanations given for depth to groundwater we found positive significant correlations for damping sensitive features (RR (0.26), P52 (0.19), SB (0.07), SDdiff (0.05)). Further it seems plausible that weak recharge signals correlate with important features for smoother hydrographs (LRec (-0.15), HPD (-0.14)). In agreement spatial recharge data, we found that clusters showing mainly smooth hydrographs with lower variability (1, 5, 8, 10, 16) are connected to lower recharge in the northern URG; clusters showing higher annual periodicity and variability and which occur mainly in the middle part of the URG (2, 4, 6) are connected to higher recharge. Nonetheless, due to missing data for France, these relations must be considered somewhat carefully.

For most of the area east of the Rhine River (Baden-Württemberg), we explored connections to the hydraulic conductivity within the uppermost aquifer (K-values, GP3, Fig. 1b) (LGRB 2007). Due to the spatially limited data, no meaningful correlation can be made with clusters, however, still a reasonable number of wells (828) can be assigned to a specific K-value. Categorical correlation analysis (Spearman) with features yields positively correlations for Skew (0.24) and SDdiff (0.18) probably because high conductivities can be

found mainly close to the Rhine River. Similarly, Jumps (0.20) are probably often caused by anthropogenic influences (abstractions, ship locks), which in turn occur preferentially in regions of high conductivities. Other correlations implicate that smoother hydrographs (HPD (-0.34)), long descending hydrograph parts (LRec (-0.23)), boundedness preferentially to an upper bound (Med01 (-0.21)), as well as the yearly maximum during spring (SB (-0.18)) seem to be related to lower hydraulic conductivities for this subset of wells. This might sound counter-intuitive, since flashy behavior is often linked to lower hydraulic conductivities, however, the main reason for flashy behavior in this area is probably the influence of the Rhine River, where high conductivities occur.

The influence of streamflows (DF2, Pr1b/4/5, Fig. 1b) was explored as the general relationships between the distance to the Rhine River and feature values. The results confirm the relation to cluster 3. Further, we found clear relationships for clusters 7 and 9. Clusters 6 and 15 showed a weaker connection, but all of the mentioned clusters show a clear spatial relation to the Rhine River. Nevertheless, they show different dynamics, which maintains the reasonability of the results. Clusters 3, 6, and 7 are closely related but flashiness of the hydrographs is decreasing from one to the other. Cluster 9 shows less periodicity than cluster 3, but both are visually similar and match for major dynamic peaks. It remains an open question what causes the different dynamics close to the streamflow. Also smaller streamflows seem to have a significant influence on groundwater at least in the southern part of our test area (Longuevergne et al. 2007). Hence, we performed a detailed streamflow distance analysis based on the Strahler classes of all streams (Text S5) in the area, derived from the Copernicus EU-HYDRO Dataset (EEA 2017). We obtained similar findings, but found a much stronger influence for cluster 15 and also a slight influence of streamflows on cluster 12.

For most conducted analyses, the correlation values are significant but rather low. This illustrates that there are distinct relations but at the same time also a lot of interactions between the influences. Correlation is nevertheless a good indicator and shows that the features express important properties of the hydrographs and thus are well selected. On the other hand, low correlation also shows, that a dynamic-based clustering is even more important, because simply grouping wells according to external factors is clearly not sufficient. Supplement Table S4 and Fig. S66 show a comprehensive overview of all explored correlations (R-values and significance).

5 Summary and Conclusions

In this work, we present the results of a newly developed semi-automated groundwater hydrograph clustering framework. We group hydrographs based exclusively on their dynamics by describing them with features specifically designed for important dynamic aspects of groundwater hydrographs. Hence, heterogeneous input data can be used, which we confirmed by a high robustness for most of our features especially towards data gaps. The combination of DS2L-algorithm with SOM allows automatic determination of the cluster number and great flexibility in terms of cluster size. It further allows the user to determine the level of detail of the clustering result. The application of two SOM-Ensembles helps to remove arbitrariness from the feature selection process, which is also a common issue in feature-based clustering. In addition, it allows to obtain robust and practice-oriented results even for groundwater observation networks that are subject to change over time. The

combination of these methods therefore creates a solid clustering framework with advantages in terms of (i) making use of heterogeneous data, (ii) operating in a comparatively highly automated manner, still leaving possibilities to adapt to specific dataset characteristics and analysis goals, as well as (iii) obtaining robust, practice-oriented results. By exchanging the describing features, the presented framework is easily transferable to other time series-clustering applications in various domains. For cluster ordering and visualization, we propose the use of a weighted correlation measure (R_w).

The clustering results illustrate the above characteristics well. Similar dynamic patterns are derived from a large data set, which can be used for further processing (e.g. forecasting) and interpretation. Our results also show, that the frequently made assumption that nearby wells have a more similar dynamic than wells further apart is only partly true, even for wells in the same aquifer. Moreover, in some cases, there are similar dynamic patterns with no clear spatial reference, which makes it important to cluster wells according to their dynamics, rather than according to spatial proximity or common aquifer properties.

We confirmed that groundwater dynamics are a complicated interaction of most diverse factors, where some of them are hard to determine or are even poorly understood at all. This makes it usually very difficult to disentangle the contributions, not to mention the mostly incomplete information on such metadata. We mainly focused on framework development, motivated by the superior goal of selecting representatives for forecasting purposes, which is why it only lies partly within the scope of this work to improve the understanding of the different factors contributing to groundwater dynamics. Thus, we have comparatively small or almost no variation in geological conditions, aquifer type and similar parameters, which is not the best starting point for a search for such correlations. Nevertheless, we hope, that our approach can contribute to this general question, besides the improved system knowledge on a local scale, which a hydrograph grouping itself already provides. This applies especially because studies of groundwater dynamics and their connections to relevant driving forces are comparatively rare yet (Giese et al. 2020). To fully exploit the potential of this method in contributing to the improvement of system knowledge, comprehensive data sets of potential influencing factors covering the complete study area should be available. The goal should be to link driving forces directly to features or indices. For this purpose more systems should be subject to research studies to explore many different characteristics and system properties. We also presume that once a better understanding of dynamic-controlling factors is in place, a prediction of ungauged locations may be possible.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11269-021-03006-y>.

Acknowledgements We thank Michel Wingerling (LUBW) for insightful assessments and discussions and Guénaél Cabanes for kindly providing the scripts of his algorithms and the permission to republish them.

Author Contributions A.W.: Conceptualization, Methodology, Software, Formal analysis, Validation, Investigation, Visualization, Writing – original draft preparation. T.L.: Conceptualization, Methodology, Writing – review and editing, Supervision. S.B.: Methodology, Writing – review and editing

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

Data Availability All groundwater data is available online from: BRGM (2018); HLNUG (2019); LUBW (2018); MUEEF (2018)

Code Availability All scripts necessary to reproduce this work are available on GitHub (DOI: 10.5281/ZENODO.3991369): www.github.com/AndreasWunsch/Groundwater-Dynamic-Clustering

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abrahart RJ, See L (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrol Process* 14(11–12):2157–2172
- Baker DB, Richards RP, Loftus TT, Kramer JW (2004) A new flashiness index: Characteristics and applications to midwestern rivers and streams. *J Am Water Resour Assoc* 40(2):503–522
- Balakrishnan PV, Cooper MC, Jacob VS, Lewis PA (1994) A study of the classification capabilities of neural networks using unsupervised learning: A comparison with K-means clustering. *Psychometrika* 59(4):509–525
- BGR (2019) Mean annual groundwater recharge of germany 1:1,000,000 (GWN1000). <https://www.bgr.bund.de/had>
- BRGM (2018) BDLISA. <https://bdlisa.eaufrance.fr/>
- Cabanes G, Bennani Y, Fresneau D (2012) Enriched topological learning for cluster detection and visualization. *Neural Netw* 32:186–195
- Caiado J, Maharaj EA, D'urso P (2015) Time-series clustering. In *Handbook of Cluster Analysis*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Boca Raton London New York, pp. 241–264
- Chang F-J, Chang L-C, Huang C-W, Kao I-F (2016) Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *J Hydrol* 541:965–976
- Chang L-C, Shen H-Y, Chang F-J (2014) Regional flood inundation nowcast using hybrid SOM and dynamic neural networks. *J Hydrol* 519:476–489
- Chen L-H, Chen C-T, Pan Y-G (2010) Groundwater level prediction using SOM-RBFN multisite model. *J Hydrol Eng* 15(8):624–631
- Chen Y, Qin B, Liu T, Liu Y, Li S (2010) The comparison of SOM and K-means for text clustering. *CIS* 3(2):268
- EEA (2017) EU-Hydro — Copernicus Land Monitoring Service. <https://land.copernicus.eu/user-corner/publications/eu-hydro-flyer>
- Geyer OF, Gwinner MP, Geyer M, Nitsch E, Simon T, Ellwanger D (2011) *Geologie von Baden-Württemberg, 5. völlig neu bearb. Aufl.* Schweizerbart, Stuttgart
- Gholami V, Khaleghi MR, Pirasteh S, Booij MJ (2021) Comparison of self-organizing map, artificial neural network, and co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial intelligence. *Water Resour Manage*
- Giese M, Haaf E, Heudorfer B, Barthel R (2020) Comparative hydrogeology - reference analysis of groundwater dynamics from neighbouring observation wells. *Hydrol Sci J* pp. 1–22
- Haaf E, Barthel R (2018) An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs. *J Hydrol* 559:222–237
- Han J-C, Huang Y, Li Z, Zhao C, Cheng G, Huang P (2016) Groundwater level prediction using a SOM-aided stepwise cluster inference model. *J Environ Manag* 182:308–321
- He J, Tan A-H, Tan C-L, Sung S-Y (2004) *On Quantitative Evaluation of Clustering Systems*, vol 11. Springer, US, Boston, MA, pp 105–133

- Heudorfer B, Haaf E, Stahl K, Barthel R (2019) Index-based characterization and quantification of groundwater dynamics. *Water Resour Res* 55(7):5575–5592
- HLNUG (2019) GruSchu. <http://gruschu.hessen.de>
- Kiang MY, Hu MY, Fisher DM (2006) An extended self-organizing map network for market segmentation—a telecommunication example. *Decis Support Syst* 42(1):36–47
- Kohonen T (2014) Matlab implementations and applications of the self-organizing map. Helsinki
- Kumar UA, Dhaniya Y (2010) Comparative analysis of SOM neural network with K-means clustering algorithm. In 2010 IEEE International Conference on Management of Innovation Technology, pp. 55–59
- LGRB (2007) Hydrogeologischer Bau und Aquifereigenschaften der Lockergesteine im Oberrheingraben (Baden-Württemberg)
- Lin G-F, Chen L-H (2005) Time series forecasting by combining the radial basis function network and the self-organizing map. *Hydrol Process* 19(10):1925–1937
- Longuevergne L, Florsch N, Elsass P (2007) Extracting coherent regional information from local measurements with Karhunen-Loève transform: Case study of an alluvial aquifer (Rhine valley, France and Germany). *Water Resour Res* 43:4
- LUBW (2006) Hydrogeologischer Bau und hydraulische Eigenschaften - 9INTERREG III A-Projekt MoNit “Modellierung der Grundwasserbelastung durch Nitrat im Oberrheingraben” / Structure hydrogéologique et caractéristiques hydrauliques - 9INTERREG III A : MoNit “Modélisation de la pollution des eaux souterraines par les nitrates dans la vallée du Rhin Supérieur”. Tech. rep., LUBW
- LUBW (2018) UDO - Umwelt-Daten und -Karten Online. <https://udo.lubw.baden-wuerttemberg.de/public/>
- Machiwal D, Singh PK (2015) Understanding factors influencing groundwater levels in hard-rock aquifer systems by using multivariate statistical techniques. *Environ Earth Sci* 74(7):5639–5652
- Mangiameli P, Chen SK, West D (1996) A comparison of SOM neural network and hierarchical clustering methods. *Eur J Oper Res* 93(2):402–417
- McMillan H, Westerberg I, Branger F (2017) Five guidelines for selecting hydrological signatures. *Hydrol Process* 31(26):4757–4761
- Melo Riveros NA, Cardenas Espitia BA, Aparicio Pico LE (2019) Comparison between K-means and Self-Organizing Maps algorithms used for diagnosis spinal column patients. *Informatics in Medicine Unlocked* 16:100206
- Mingoti SA, Lima JO (2006) Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *Eur J Oper Res* 174(3):1742–1759
- Moradkhani H, Hsu K-L, Gupta HV, Sorooshian S (2004) Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. *J Hydrol* 295(1):246–262
- MUEEF (2018) Geoportel Wasser. <http://geoportel-wasser.rlp.de/servlet/is/8183/>
- Naranjo-Fernández N, Guardiola-Albert C, Aguilera H, Serrano-Hidalgo C, Montero-González E (2020) Clustering groundwater level time series of the exploited almonte-marismas aquifer in southwest Spain. *Water* 12(4):1063
- Nourani V, Alami MT, Voutsoughi FD (2015) Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *J Hydrol* 524:255–269
- Région Alsace - Strasbourg (1999) Bestandsaufnahme der Grundwasserqualität im Oberrheingraben / Inventaire de la qualité des eaux souterraines dans la vallée du Rhin Supérieur
- Richter BD, Baumgartner JV, Powell J, Braun DP (1996) A method for assessing hydrologic alteration within ecosystems. *Conserv Biol* 10(4):1163–1174
- Toth E (2009) Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. *Hydrol Earth Syst Sci* 12
- Toth E (2013) Catchment classification based on characterisation of streamflow and precipitation time series. *Hydrol Earth Syst Sci* 17(3):1149–1159
- Vesanto J (2005) SOM toolbox: Implementation of the algorithm. <http://www.cis.hut.fi/projects/somtoolbox/documentation/somalg.shtml>
- Wang X, Smith KA, Hyndman RJ (2006) Characteristic-based clustering for time series data. *Data Min Knowl Disc* 13(3):335–364

Authors and Affiliations

Andreas Wunsch¹  · Tanja Liesch¹  · Stefan Broda² 

Tanja Liesch
tanja.liesch@kit.edu

Stefan Broda
stefan.broda@bgr.de

¹ Karlsruhe Institute of Technology (KIT), Institute of Applied Geosciences, Division of Hydrogeology, Kaiserstr. 12, Karlsruhe 76131, Germany

² Federal Institute for Geosciences and Natural Resources (BGR), Wilhelmstr. 25-30, Berlin 13593, Germany