




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

A. Wunsch\*  · Liesch, T.  · Broda, S. 

DOI [10.1007/s11269-021-03006-y](https://doi.org/10.1007/s11269-021-03006-y)

GitHub [AndreasWunsch/Groundwater-Dynamic-Clustering](https://github.com/AndreasWunsch/Groundwater-Dynamic-Clustering)

## Content

1. Text S1 to S5
2. Figures S1 to S66
3. Tables S1 to S4

**Summary** This document serves as electronic supplementary to the article: Feature-based Groundwater Hydrograph Clustering using unsupervised Self-Organizing-Map-Ensembles. It summarizes additional information on several topics. **Text S1** describes and discusses influences on groundwater dynamics with a special regard to the study area. All abbreviations refer to Figure 1b of the main text. **Text S2** gives information on calculation formulas and additional explanations on all self-designed features with respect to Table 1 of the manuscript. **Text S3** focuses on a detailed description of the workflow referring to Section 3.3 and Figure 2 of the main text. **Text S4** and **Table S1** summarize the results of all experiments conducted to explore the robustness of our features against data gaps, noise and time series length. **Text S5** explains the streamflow influence analysis based on Strahler classes.

Texts S1 to S5 are followed by graphics showing detailed results of the visual skill test results of all features used in this study (**Figures S1 to S13**). The detailed results of the feature robustness experiments as a supplement to Text S4 and Table S1 are given in **Figures S14 to S26** and **Tables S2 to S3**. **Figure S27** gives detailed information on correlation between all features. **Figure S28** shows the feature values as boxplots for the final clusters, **Figures S29 to S65** show detailed results of our clustering in the Upper Rhine Graben (stacked hydrograph plots and maps of respective well positions). **Table S4** gives a summary of the correlation analysis results between features and influencing factors, **Figure S66** shows according scatter plots.

## Text S1. Groundwater dynamics and its influences in the upper Rhine graben

This section refers to Section 2 (Data and Study area) of the main text and especially describes and discusses the factors mentioned in Figure 1b.

One of the main processes with influence on groundwater dynamics in the region is groundwater recharge (Pr1), either directly (Pr1a/b) or as inter-aquifer exchange (Pr1c). Direct recharge is a highly complex process and occurs diffuse through the unsaturated zone (Pr1a) or localized (DF2/Pr1b). Recharge in general also depends on many other factors like precipitation (physical state, amount, intensity) (DF4), temperature (DF4), topography (GP1), vegetation (GP2), geology (GP3), soil moisture (DF5) etc. (e.g. Jasechko et al., 2014; Alley et al., 2002). Some of these, especially precipitation and temperature, in

\*Corresponding Author: Andreas Wunsch

Karlsruhe Institute of Technology (KIT), Institute of Applied Geosciences, Division of Hydrogeology, Kaiserstr. 12, 76131 Karlsruhe, Germany  
E-mail: andreas.wunsch@kit.edu

turn are driven by global climatic patterns (DF4), which especially in humid regions have a significant influence on groundwater levels (Cuthbert, 2014) and generally influence factors like land-use and vegetation (GP2) directly. These in turn have a strong impact on soil-moisture (DF5) and evapotranspiration (Pr2). Further, mainly during long dry seasons, shallow groundwater is exposed to the risk of strong direct groundwater evaporation (Balugani et al., 2017) even in moderate climate with significant portions, as shown by Lam et al. (2011). The URG is one of the warmest areas in Germany and the yearly precipitation within the Graben is in the order of 500 to 900 mm per year, the adjacent mountain regions can reach cumulative rainfalls of 2000 mm per year (Thierion et al., 2012). Mean annual groundwater recharge in our dataset covers a range from 0 mm (mainly floodplains of the Rhine) to about 350 mm/a, with a mean value of about 150 mm/a. In general, the diffuse recharge in the northern part is comparably low, while the highest recharge values mostly occur in the middle URG between the cities of Offenburg and Rastatt (BGR, 2019). Dominant land use types within the URG are agricultural areas of different types (37%), on par with artificial surfaces (36%), the rest are mostly forests/semi-natural surfaces (22%) (CORINE Land Cover, 2018).

Geology (GP3), thus, material properties (permeability/hydraulic conductivity, effective porosity) or more generally speaking the aquifer type (porous, fractured, karstic), also plays a major role in controlling groundwater dynamics. Porous unconsolidated gravel or sand aquifers like in the URG usually show highest matrix porosities, often going along with high hydraulic conductivity and high storage capacity. Also, the regional geological setting is of great importance, since the development of local and regional groundwater flow-systems (DF3), thus the lateral recharge (Pr1d) within an aquifer, depends on it (Toth, 2009). Confined and unconfined aquifers (GP4) are known to react differently to atmospheric pressure changes or groundwater withdrawal (Alley et al., 2002; Hölting and Coldewey, 2013). The mean depth to groundwater (GP5) is also an important factor concerning groundwater dynamics as the recharge signal is increasingly damped with depth (Pr3), filtering seasonal variation patterns leaving only multi-annual periodicities. Overlying layers with lower hydraulic conductivities can amplify this low-pass filter effect (e.g. Corona et al., 2018). The study area comprises mainly unconfined sand/gravel aquifers of generally high storage coefficients and high hydraulic conductivities in the order of  $10E-4$  to  $10E-3$  m/s (LUBW, 2006). Hydrographs used in this study are from the uppermost aquifer, with very shallow mean depths to groundwater ( $<5$  m bgl for 70% of the wells), rising to a maximum of about 20-30 m towards the Graben edges. A rather shallow gradient towards the north of the Graben and at the same time from the Graben edges towards the graben center controls the regional groundwater flow-systems (Thierion et al., 2012). Towards the Graben edge, local inflow from adjusting fissured aquifers or alluvial fans from side-valleys may dominate the flow regime and result in steeper gradients, towards the Rhine River as the main receiving streamflow of the region.

Surface water interactions (DF2), already mentioned as a source of local recharge, are usually important driving forces of groundwater dynamics. Important processes and driving forces in this context are for example streamflow in- and exfiltration (Pr4), bank storage (Pr5), tides, waves, as well as floods (DF2a)(Alley et al., 2002; Cloutier et al., 2014). In the study area, the main surface water body is the Rhine River, with a strong influence on groundwater dynamics, up to several hundreds of meters in distance. To a lesser degree, there are also smaller streams from the adjacent mountain ranges that strongly affect groundwater dynamics on local scale (Longuevergne et al., 2007). Besides natural interaction, especially in floodplains and along the ancient river course, anthropogenic intervention like correction of the streambed course or weir locks and dams influence the dynamics in many parts along the streams.

Anthropogenic actions in general cannot only influence streamflows but also strongly alter groundwater dynamics directly (Stoll et al., 2011). Typical influences in general, also widely present in the study area, are land-use changes over wide areas, landscape-engineering actions (e.g. river course modifications and dredging lakes), recharge inhibition by surface sealing in urban areas, abstraction for drinking water supply or industrial purposes, artificial infiltration, and irrigation in agricultural areas, which increased in the study area particularly in recent years. Especially direct groundwater interactions like abstractions and infiltrations (DF1) are most important because on local scale pumping patterns can partly or even completely superimpose the natural groundwater dynamics. Especially in the northern part of the URG intensive groundwater management is applied by managing extraction rates and artificial aquifer recharge. Besides the increasing water demand in these areas this is especially necessary to protect ecosystems and infrastructure from land-subsidence and groundwater-floodings (Bouwer, 2002; Regierungspräsidium Darmstadt, 1999).



## Text S2. Self-designed Features: Calculation and Explanations

In the following, we introduce detailed information on the interpretation and calculation of the self-designed features. For a description of features derived from literature, please refer to the respective publications. Features based on standard statistics are not explained in detail either.

### Feature: RangeRatio

The feature Range Ratio (RR) considers the ratio of the mean annual range ( $\overline{rng}_y$ ) to the maximum overall range ( $rng_{max}$ ) of the groundwater time series. Primary purpose of this feature is to differentiate between hydrographs with and without superimposing long-periodic signals. In addition to such periodicities, periods of increased or decreased groundwater levels as well as outliers or partial outliers can also lead to low range ratio values. For the calculation of  $\overline{rng}_y$ , only years with a maximum of 16 missing values are used. This ensures a realistic range per year. Ranges are calculated on original, unscaled, unnormalized hydrographs.

$$RR = \frac{\overline{rng}_y}{rng_{max}} \quad (1)$$

### Feature: Periodicity

The feature Periodicity (P52) is designed as a measure of the strength of the annual cycle of a hydrograph ( $TS$ ). For this purpose, the mean annual periodicity ( $TS_{periodic}$ ) is extracted by simply averaging the corresponding values of the individual years. For this purpose, a period length of 52 (number of weeks/measured values per year) is assumed and all values are averaged at an interval of 52. The final feature value is obtained by calculating the Pearson R between  $TS$  and  $TS_{periodic}$ . P52 is calculated on original, unscaled, unnormalized hydrographs.

$$P52 = corr(TS, TS_{periodic}) \quad (2)$$

### Feature: $SD_{diff}$

The feature  $SD_{diff}$  describes how often strong rates of changes within a time series  $TS$  occur. It is therefore a measure of flashiness and variability. We use the standard deviation  $\sigma$  of the first derivative of the original, unscaled, unnormalized time series data for calculation:

$$SD_{diff} = \sigma\left(\frac{d}{dt}TS(t)\right) \quad (3)$$

### Feature: Longest Recession

The feature Longest Recession (LRec) searches for the longest section of a time series  $TS$  without rising groundwater levels. This allows to identify time series with sections with (unnaturally) long falling groundwater levels, where even the annual cycle can be completely lost, as well as to group together smoother time series.

$$LRec = \left(\frac{d}{dt}TS(t), \frac{d}{dt}TS(t+1), \dots\right)_{max}, \quad \forall \frac{d}{dt}TS(t) \leq 0 \quad (4)$$

### Feature: Jumps

The feature Jumps is designed to detect inhomogeneities/breaks resulting in jumps in the time series. For this purpose, the absolute change of the mean value of successive years is considered. Analogous to the feature RR, only years with a maximum of 16 missing values are included, to ensure that the mean value realistically represents the annual variation. In contrast to the previously mentioned features, the calculation of jumps is based on z-scored data, since inhomogeneities are to be considered independent of the scale and only relative to the dynamics of the individual time series. The standardization of the

maximum mean value change by the mean change of successive years makes it possible to distinguish between true inhomogeneities and regularly occurring changes. To highlight large jumps values, the values are finally squared. For  $J$  as the time series of the annual mean values, Jumps is calculated as follows:

$$Jumps = \left( \frac{|\frac{d}{dt}J(t)|_{max}}{|\frac{d}{dt}J(t)|} \right)^2 \quad (5)$$

*Feature: Seasonal Behaviour*

The feature Seasonal Behaviour (SB) is used to differentiate the time series according to the position of their yearly maximum and minimum. For this purpose, all corresponding values of the time series are averaged monthly and a mean annual cycle  $AC$  of the monitoring well consisting of twelve values is generated.  $AC$  is compared (Pearson correlation) with the expected seasonality for this area (model curve  $MC$ ): sine curve with maximum in March and minimum in September). In order to also include the normalized amplitude, the euclidean distance (L2 norm) of both curves  $AC$  and  $MC$  is calculated. If the correlation of both curves is smaller than zero, the L2 norm between  $AC$  and the inverse model curve  $MC_{inv}$  (minimum and maximum exchanged) is used. SB is based on z-scored data to enable a meaningful comparison of  $AC$  and  $MC$ .

$$R = corr(AC, MC) \quad (6)$$

$$SB = \frac{R}{d} \{ d = \|AC - MC\|, \quad \forall R \geq 0 \quad d = \|AC - MC_{inv}\|, \quad \forall R < 0 \} \quad (7)$$

*Feature: Yearly Variance*

The feature Yearly Variance ( $Y_{var}$ ) is calculated as the median of the annual variances  $\mu$  of a z-scored time series. Again, the calculation is made only for years  $J$  with less than 17 missing values.  $Y_{var}$  captures a combination of periodicity and variance of a time series. The correlation to the periodicity feature P52 is often high, since time series with strong annual cycle usually also show high annual variances on average. Conversely, however, high  $Y_{var}$  values can also be achieved if no strong regular annual periodicity is present.

$$Y_{var} = \widetilde{\mu}(J) \quad (8)$$

### **Text S3. Detailed description of workflow**

Figure 3 of the main text summarizes the workflow applied in this study. in the following we extent section 3.3 to share more insights on the applied methodology.

First, depending on data characteristics and hydrogeological conditions, an adequate set of features with good explanatory power is selected. Next, a gradual optimization of the cluster parameters follows (optionally including a revision of step one). From experience we conclude that both steps usually require incremental adjustment and repetition to obtain a promising feature-parameter combination. They are hereby extremely valuable to create and include existing expert knowledge in terms of which features might be important and how many features will be approximately needed to describe major dynamics aspects. The result is processed by the first ensemble calculation with the goal of finding a most suitable feature combination.

Like all ANNs, SOM are sensitive to the initialization procedure. Despite we use linear initialization (Vesanto, 2000) to minimize this effect, still, some dependency on the input-order remains. This makes it necessary to also test feature-order permutations, usually by applying a grid search for fewer feature numbers. Unfortunately, this causes factorial growth of the computational complexity. However, the number of calculations generally can be reduced significantly by deducing constraints from prior knowledge (e.g. some features must be included, min./max number of features etc.), which normally can already be derived from steps one and two of the workflow. Another way to handle this problem might be to partly ignore the influence of the feature order and test only permutations of promising feature

combinations, since, in our experience, the latter usually seems to be more important. There is also a certain entanglement between feature number and cluster parameters, meaning that the cluster parameters are already matched to an order of magnitude in terms of the number of features. The ensemble is therefore not completely free of assumptions, which is certainly desirable for the aforementioned reasons. We combined five different internal validation indices to judge cluster quality (Caliński-Harabasz criterion (CH) (Caliński and Harabasz, 1974), McClain-Rao criterion (MR) (McClain and Rao, 1975), PBM-Index (Pakhira et al., 2004), Ratkowsky-Lance criterion (RL) (Ratkowsky and Lance, 1978), C-Index (Hubert and Schultz, 1976)). The first three indices (CH, MR, PBM) use both compactness and separation of the clusters, RL judges only cluster separation, C-Index in turn only cluster-compactness. For index calculation we use the R-package ClusterCrit (Desgraupes, 2018). These validation indices determine a ranking of all members to select the supposedly most suitable realization.

Next, we use delete-d-Jackknifing resampling to build the second ensemble (10.000 members). The purpose is to simulate changes in the observational network by manipulating the input data set, and to obtain cluster results as robust as possible. Accordingly, the data is randomly reduced by  $d$  samples without replacement for each member. In agreement to Sinharay (2010) we choose the lower bound for  $d$  to be the square root of the available sample number.

For each member, cluster solutions are calculated and then combined by applying Voting-Consensus (Shestakov, 2017; Ayad and Kamel, 2010), which improves the quality and consistency/robustness of the result severely (Alqurashi and Wang, 2018; Ghosh and Acharya, 2011; Vega-Pons and Ruiz-Shulcloper, 2011). Due to very slight start value dependency of voting, we repeat this step 12 times.

The last step includes the evaluation of the cluster results by judging the clusters mathematically, visually and spatially. The workflow can be re-entered at any step to optimize the outcome.

#### **Text S4. Feature Robustness Results**

To better examine the properties and data requirements of the applied features we designed three experiments. The first two experiments try to answer the question how strongly the features react to missing values and to white noise in the data respectively. Thus, 0%-25% of each time series is randomly replaced by white noise or data gaps in 0.25% steps (50 times each) and both the absolute characteristic values and their changes compared to the initial undisturbed values are examined. In order to estimate how long a time series has to be at least to provide a representative feature value, the features for systematically varied time series lengths were calculated in experiment three. Starting from 2016, the time series length was extended in 1-year steps until 1986. For this experiment we used only a subgroup of about 50% of the data set, which had complete data over the 30 years. To make the feature values and changes comparable, in all experiments the features were standardised, using the respective mean and standard deviation from the (undisturbed) 30-year feature values.

We conducted these three experiments to explore the robustness of the 13 tested features and present the results hereafter. As reference values within each experiment the undisturbed features values (no additional gaps, noise or shortened time series) were used. We compare the changes after standardization using the mean and standard deviation from the (undisturbed) 30-year feature values. Thus, the unit is basically standard deviations. Table S1 summarizes the results of all three experiments.

We can show that most features only react little ( $<0.1$  with 25% missing values) to additional data gaps. In contrast, adding white noise leads to much higher differences much faster. Though one might think, this could lead to unstable results for noisy datasets, in reality this is probably not the case. Little noise from unknown sources is hard to recognize at all and will not lead to strong differences in feature calculation. Strong noise, however, which causes higher differences in the features, usually can be detected as outliers and removed hereafter. Therefore, data should always be carefully checked for implausible outliers in preprocessing. Experiment three shows that time series length seems to have a constant influence on the feature values. We found a steady increase of differences the shorter the time series, up to strong increases for lengths of only few years. No threshold value that might serve as recommendation as minimum length can be found, thus we rather generally conclude that the longer the time series the better. Features that are not robust and show bad performance, or cause unsatisfying cluster results, should usually be ruled out by the visual skill test or by the feature selection ensemble. Please check the supplement for detailed information on feature robustness results. Answering how

**Table S1** Median influences of data gaps, white noise and time series length on standardized feature values. The table shows the absolute values of the differences between the according disturbed values and the undisturbed values (no additional noise, data gaps, full length).

Feature	Added White Noise				Added Data Gaps			
	1%	5%	10%	25%	1%	5%	10%	25%
RR	0.13	0.49	0.75	1.14	0.00	0.01	0.02	0.04
Skew	0.00	0.03	0.05	0.12	0.00	0.00	0.00	0.01
P52	0.02	0.08	0.15	0.34	0.00	0.01	0.03	0.07
SDdiff	0.41	1.37	2.10	3.42	0.01	0.03	0.06	0.15
LRec	0.00	0.20	0.40	0.69	0.00	0.06	0.11	0.23
Jumps	0.01	0.03	0.05	0.11	0.00	0.02	0.03	0.12
SB	0.00	0.02	0.03	0.08	0.00	0.00	0.01	0.01
Med01	0.00	0.11	0.30	0.49	0.00	0.00	0.00	0.00
HPD	0.10	0.31	0.42	0.52	0.01	0.03	0.04	0.10
<i>RBI</i>	<i>0.03</i>	<i>0.24</i>	<i>0.46</i>	<i>0.99</i>	<i>0.00</i>	<i>0.01</i>	<i>0.01</i>	<i>0.03</i>
<i>Yvar</i>	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	<i>0.01</i>	<i>0.01</i>	<i>0.04</i>	<i>0.09</i>	<i>0.22</i>
<i>SEM</i>	<i>0.08</i>	<i>0.27</i>	<i>0.38</i>	<i>0.48</i>	<i>0.01</i>	<i>0.03</i>	<i>0.04</i>	<i>0.10</i>
<i>LPD</i>	<i>0.14</i>	<i>0.68</i>	<i>1.31</i>	<i>2.82</i>	<i>0.01</i>	<i>0.04</i>	<i>0.08</i>	<i>0.18</i>
	Time Series Length [years]							
	30	25	20	15	10	5	3	1
RR	0.00	0.05	0.24	0.40	0.87	1.47	2.10	NaN
Skew	0.00	0.08	0.17	0.29	0.27	0.39	0.53	0.51
P52	0.00	0.07	0.16	0.18	0.45	0.78	0.78	2.68
SDdiff	0.00	0.02	0.04	0.06	0.09	0.10	0.14	0.16
LRec	0.00	0.00	0.00	0.00	0.06	0.34	0.40	0.92
Jumps	0.00	0.14	0.23	0.44	0.91	1.39	1.76	NaN
SB	0.00	0.04	0.07	0.12	0.16	0.22	0.32	0.61
Med01	0.00	0.10	0.29	0.46	0.53	0.60	0.85	0.95
HPD	0.00	0.05	0.08	0.10	0.14	0.15	0.17	0.31
<i>RBI</i>	<i>0.00</i>	<i>0.06</i>	<i>0.15</i>	<i>0.21</i>	<i>0.51</i>	<i>0.76</i>	<i>1.38</i>	<i>NaN</i>
<i>Yvar</i>	<i>0.00</i>	<i>0.15</i>	<i>0.26</i>	<i>0.47</i>	<i>0.60</i>	<i>1.49</i>	<i>2.13</i>	<i>5.08</i>
<i>SEM</i>	<i>0.00</i>	<i>0.03</i>	<i>0.05</i>	<i>0.08</i>	<i>0.12</i>	<i>0.17</i>	<i>0.19</i>	<i>0.27</i>
<i>LPD</i>	<i>0.00</i>	<i>0.02</i>	<i>0.03</i>	<i>0.05</i>	<i>0.07</i>	<i>0.09</i>	<i>0.10</i>	<i>0.13</i>

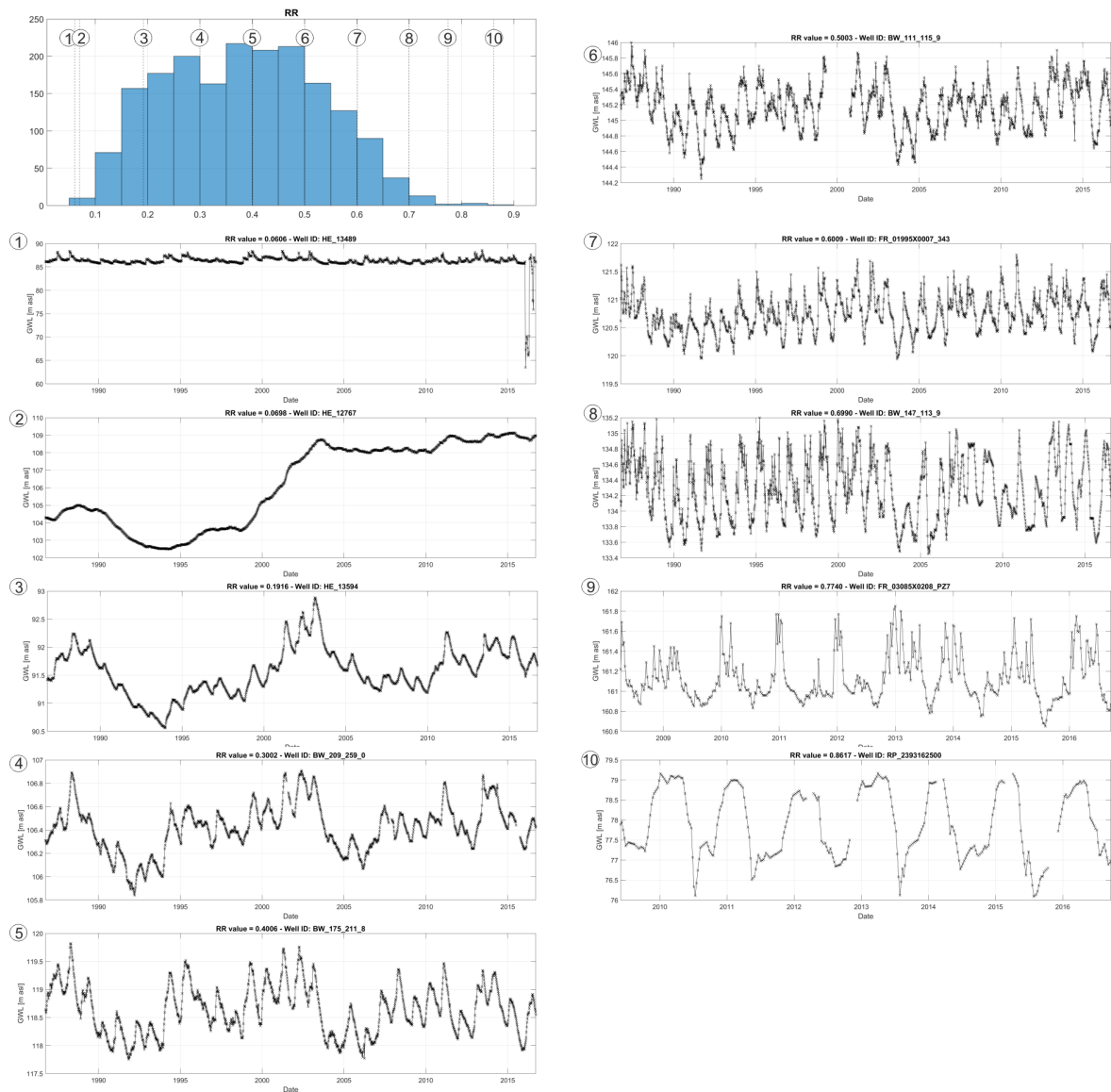
these disturbances alter the clustering result is extremely difficult, because additional factors such as the ensemble and the consensus voting approaches also influence the final results. Extensive and thorough experiments would be necessary to investigate these interactions, which is why this question lies beyond the scope of this work, but would be worth to be answered in future research.

### Text S5. Analysis of streamflow influence based on Strahler classes

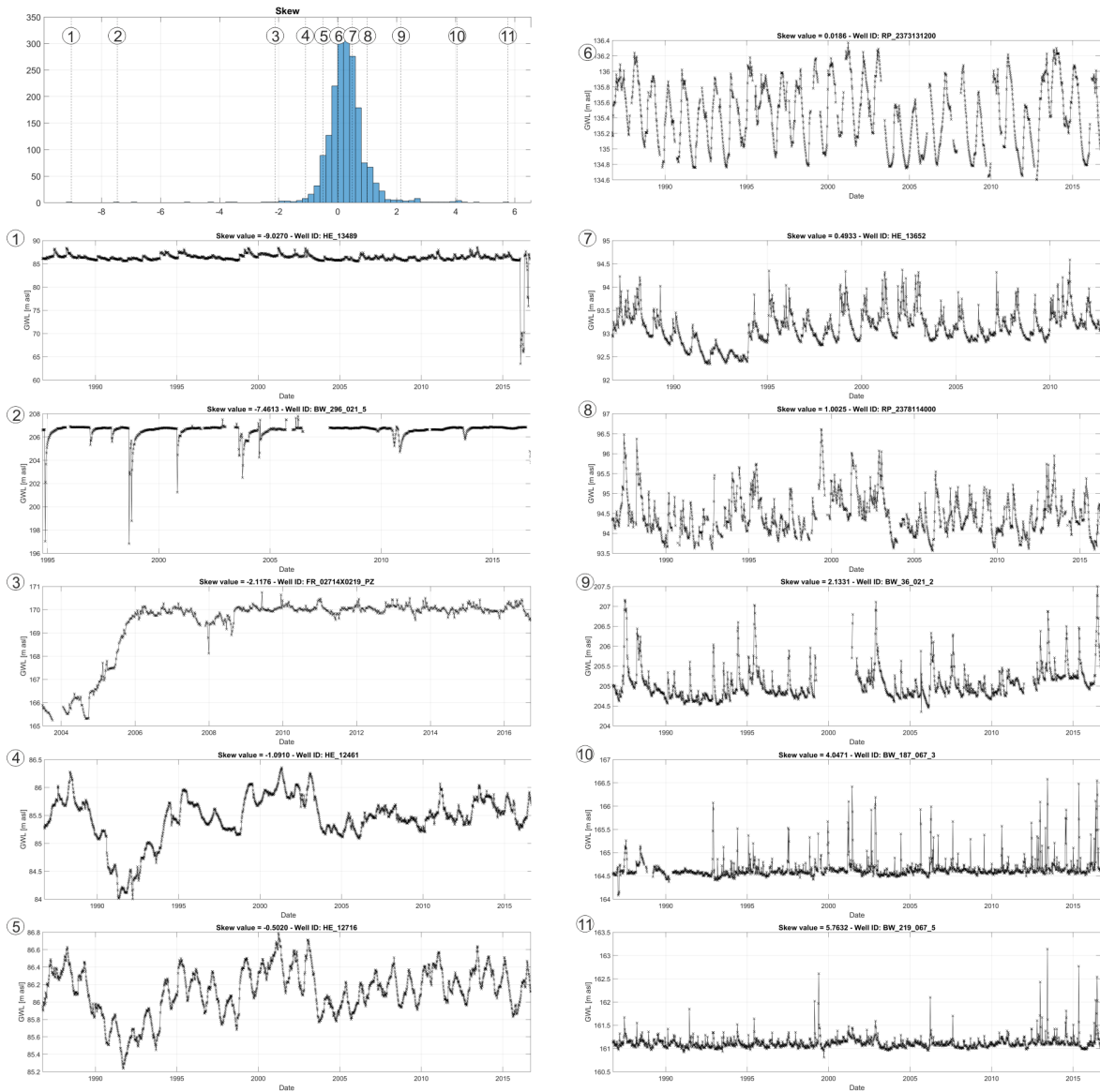
In the following, we introduce detailed information on the calculation of the streamflow analysis based on the Strahler classes of the streamflows in the study area. The degree of interaction between streamflows and groundwater depends both on the distance to the streamflow and the size of the streamflow. We calculate the streamflow influence  $SFinf$  as follows:

$$SFinf = \frac{StrahlerClass^3}{Distance} \quad (9)$$

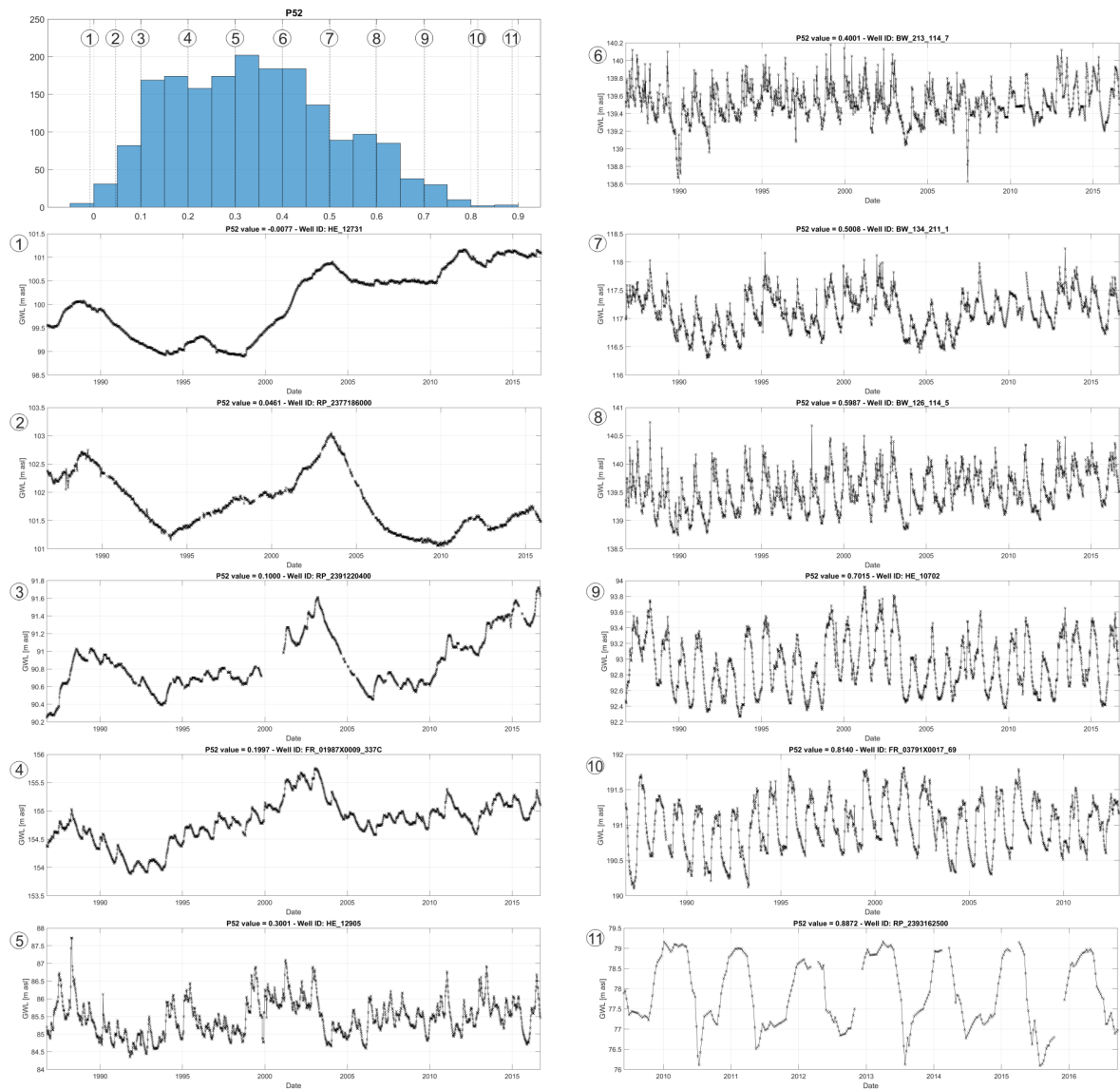
In this way we take into account that although large rivers may be further away from a particular groundwater monitoring well, the overall influence may be greater than the influence of a nearby but very small streamflow. We calculate the influence for each groundwater observation well to the nearest streamflow of the respective class and select the highest influence respectively. This also means that for wells that may not be influenced by streamflows at all, we choose the influence for the largest streamflow the Rhine River (Strahler: 7) in the region. However, this influence is then very small.



**Fig. S1** Selected results of the visual skill test that support the explanatory power of feature RR.



**Fig. S2** Selected results of the visual skill test that support the explanatory power of feature Skew.



**Fig. S3** Selected results of the visual skill test that support the explanatory power of feature P52.

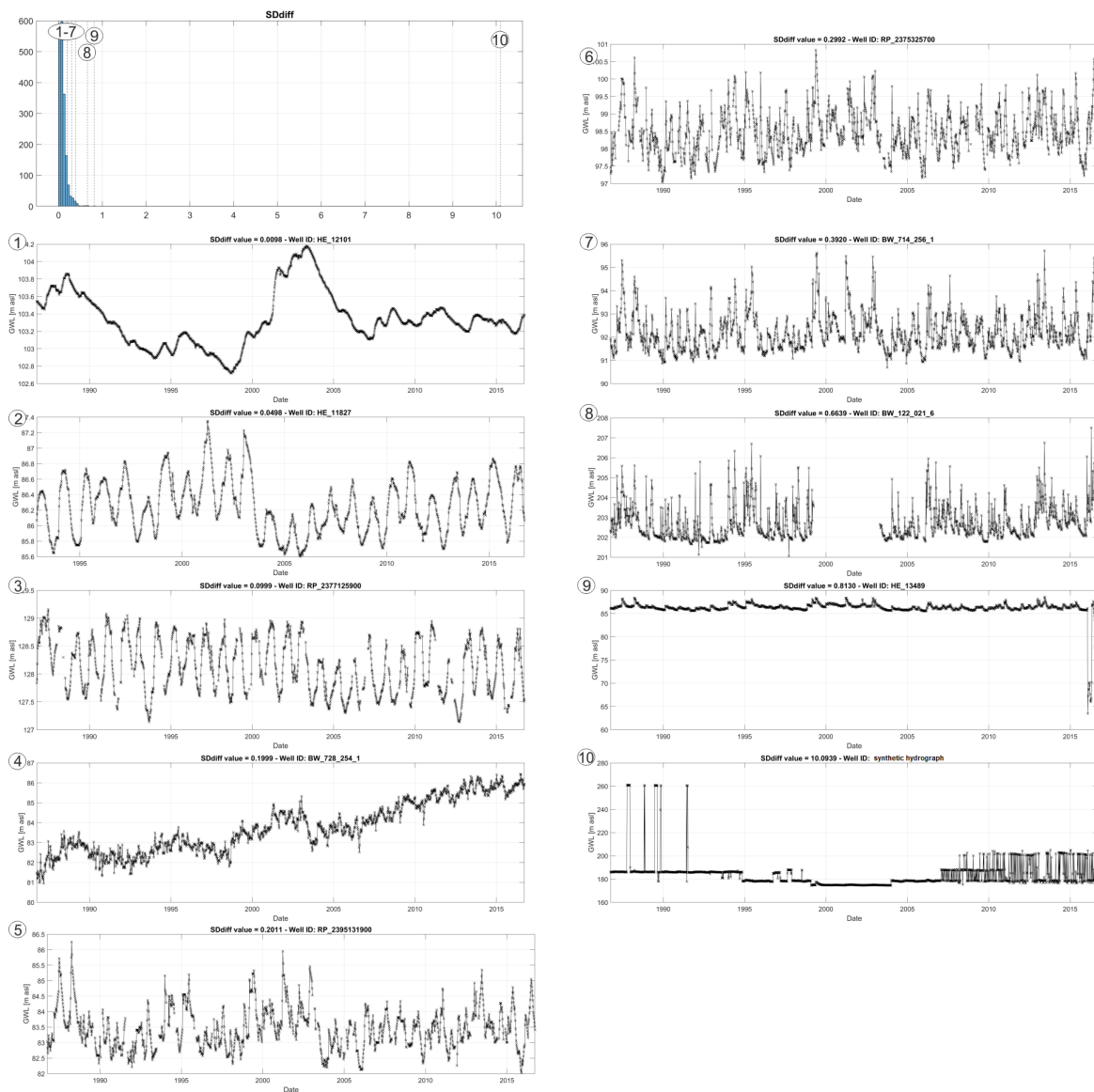
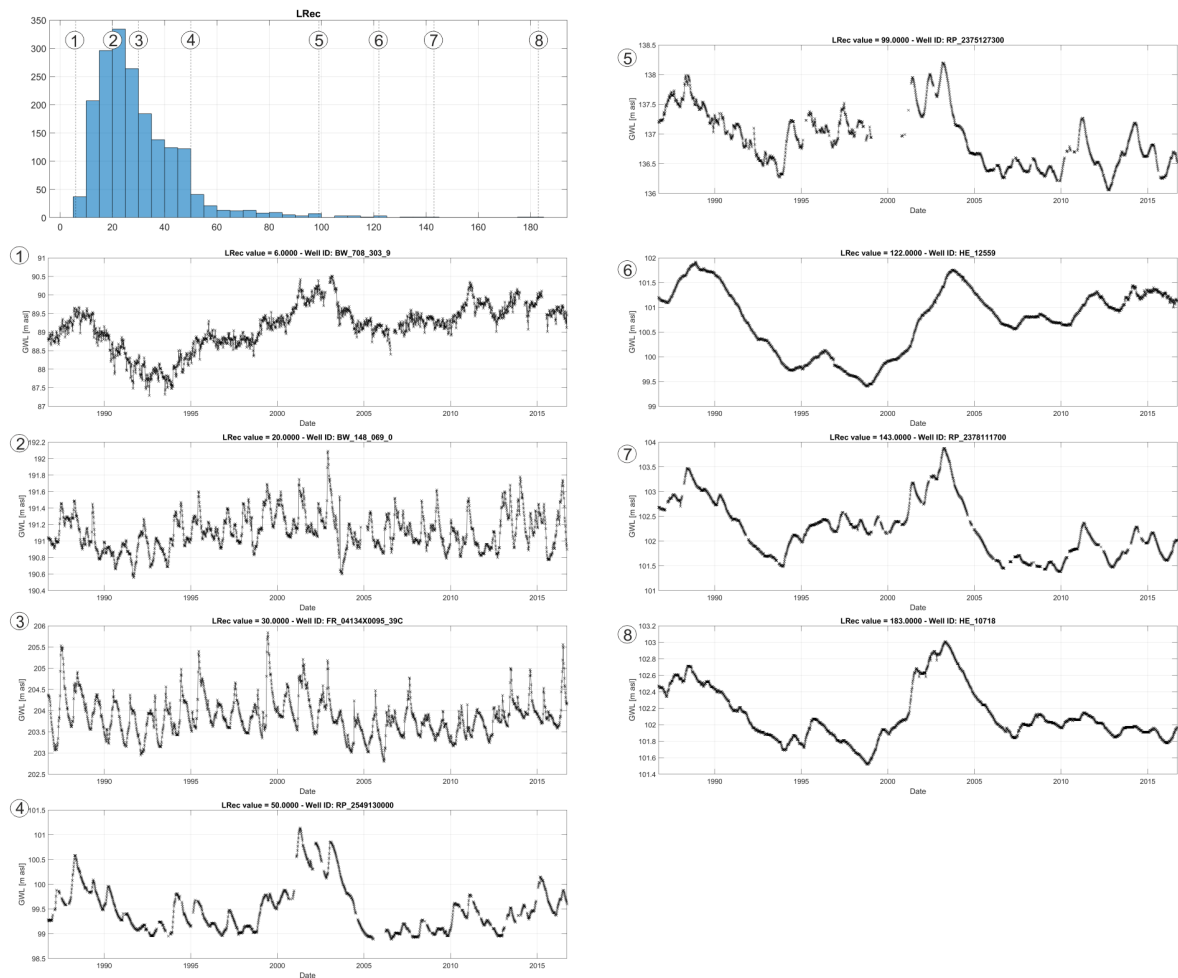
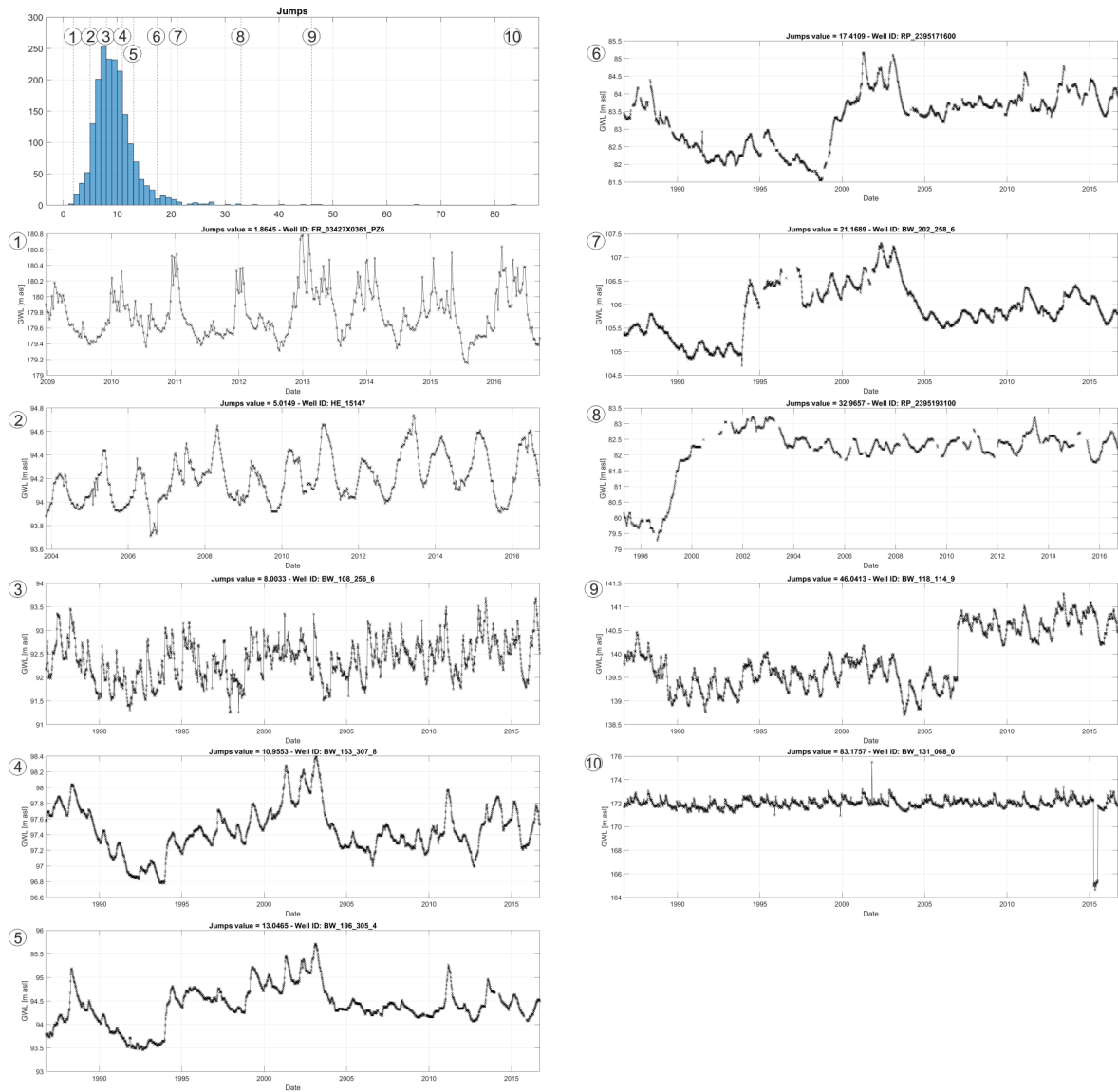


Fig. S4 Selected results of the visual skill test that support the explanatory power of feature  $SD_{diff}$ .





**Fig. S5** Selected results of the visual skill test that support the explanatory power of feature LRec.



**Fig. S6** Selected results of the visual skill test that support the explanatory power of feature Jumps.

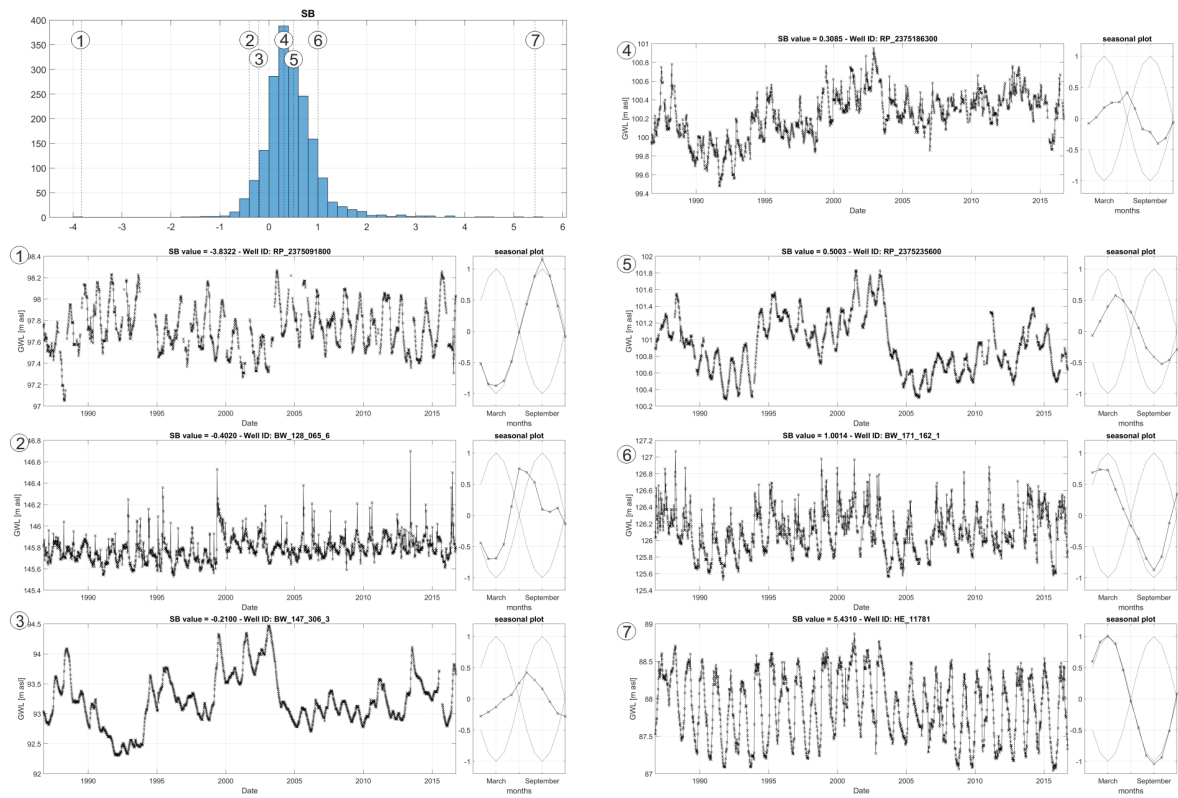


Fig. S7 Selected results of the visual skill test that support the explanatory power of feature SB.

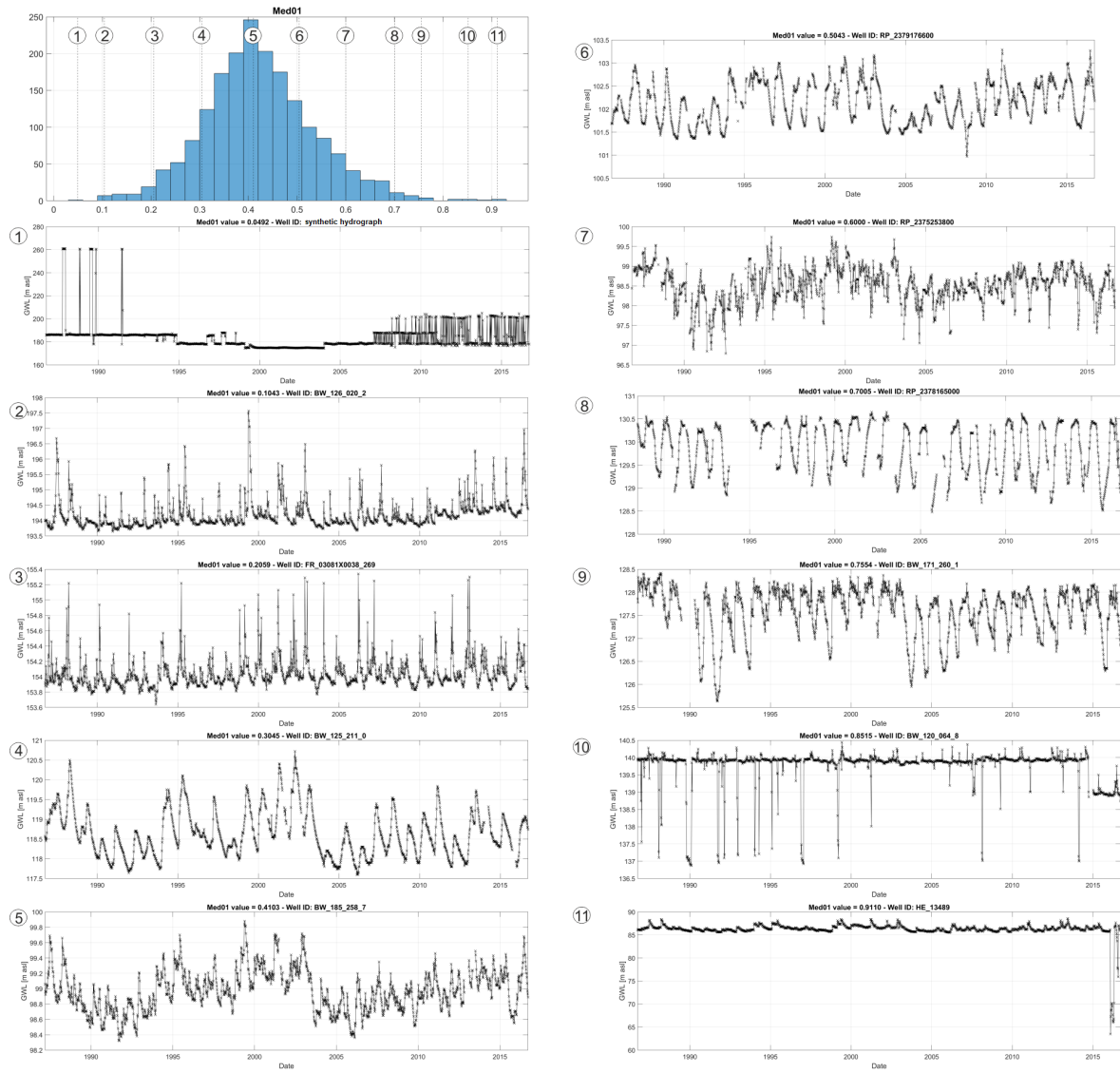
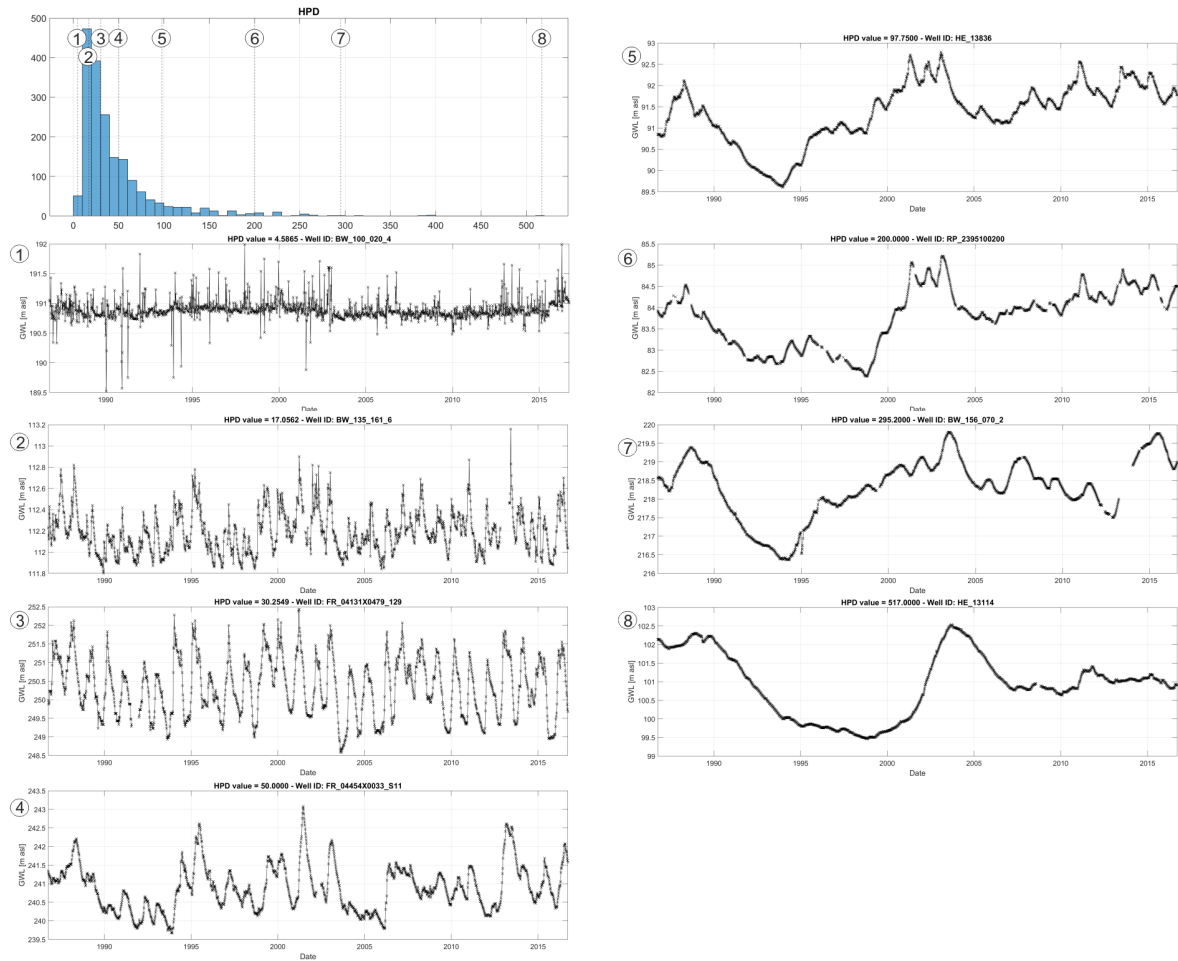


Fig. S8 Selected results of the visual skill test that support the explanatory power of feature Med01.



**Fig. S9** Selected results of the visual skill test that support the explanatory power of feature HPD.

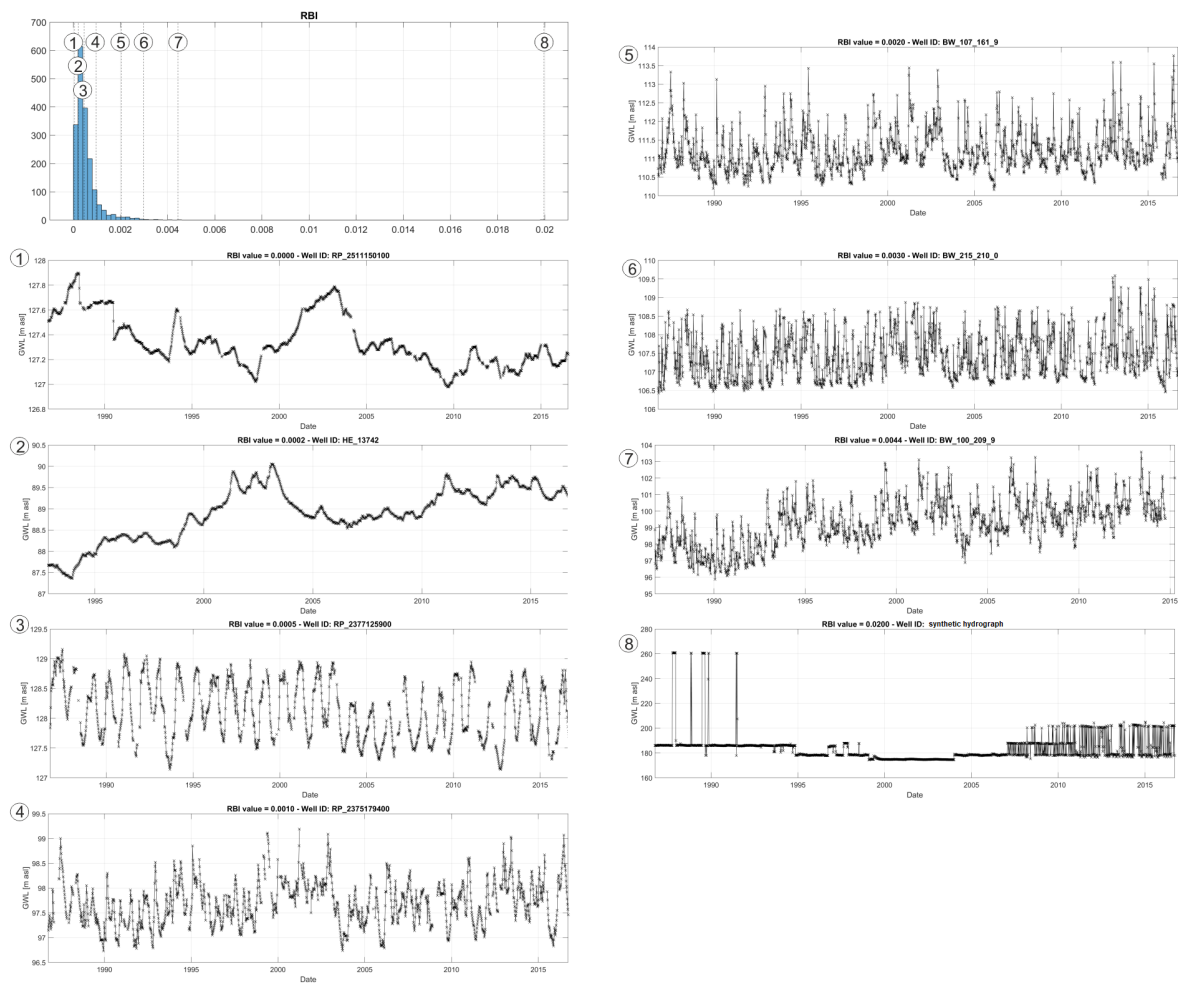


Fig. S10 Selected results of the visual skill test that support the explanatory power of feature RBI.

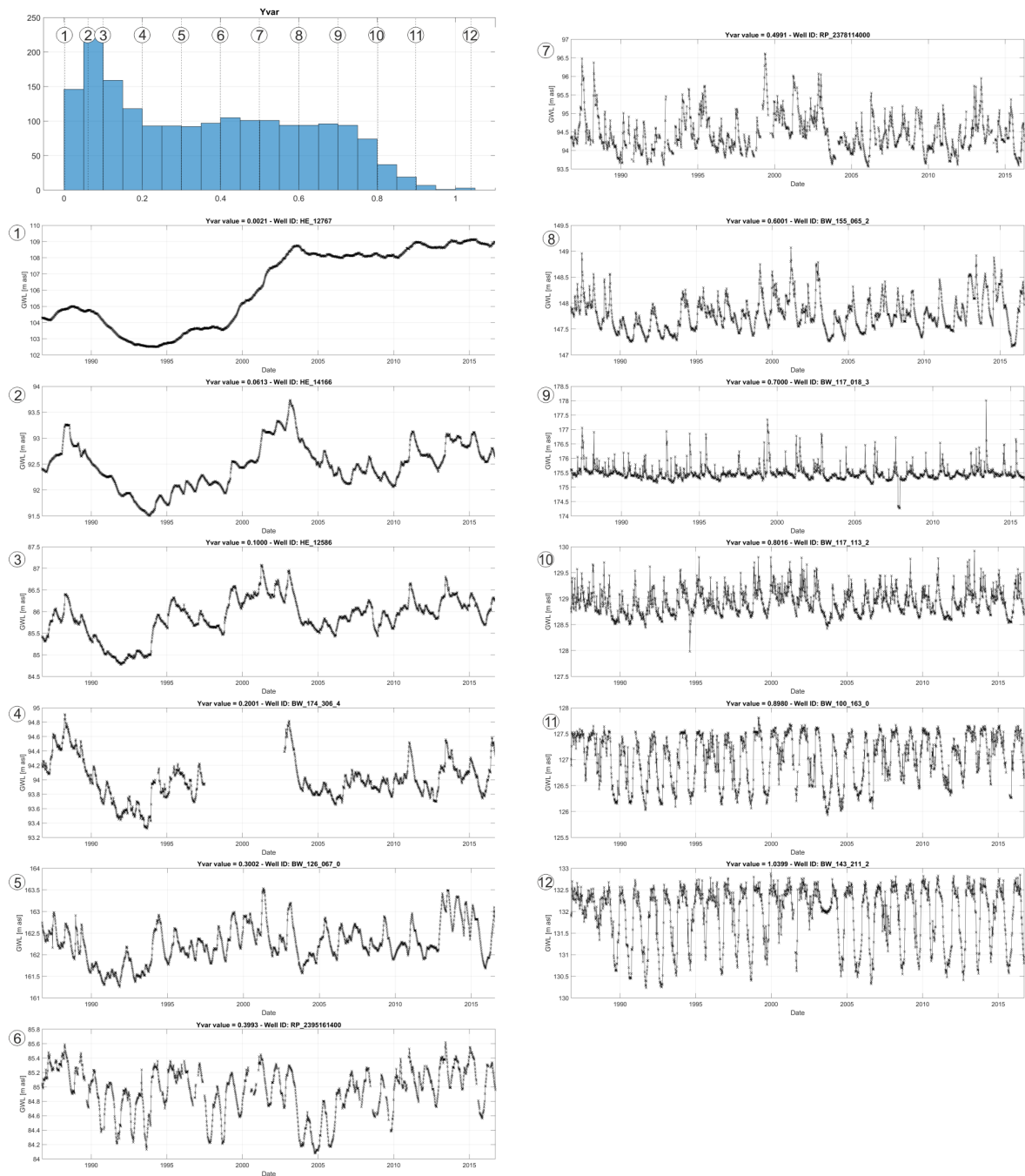
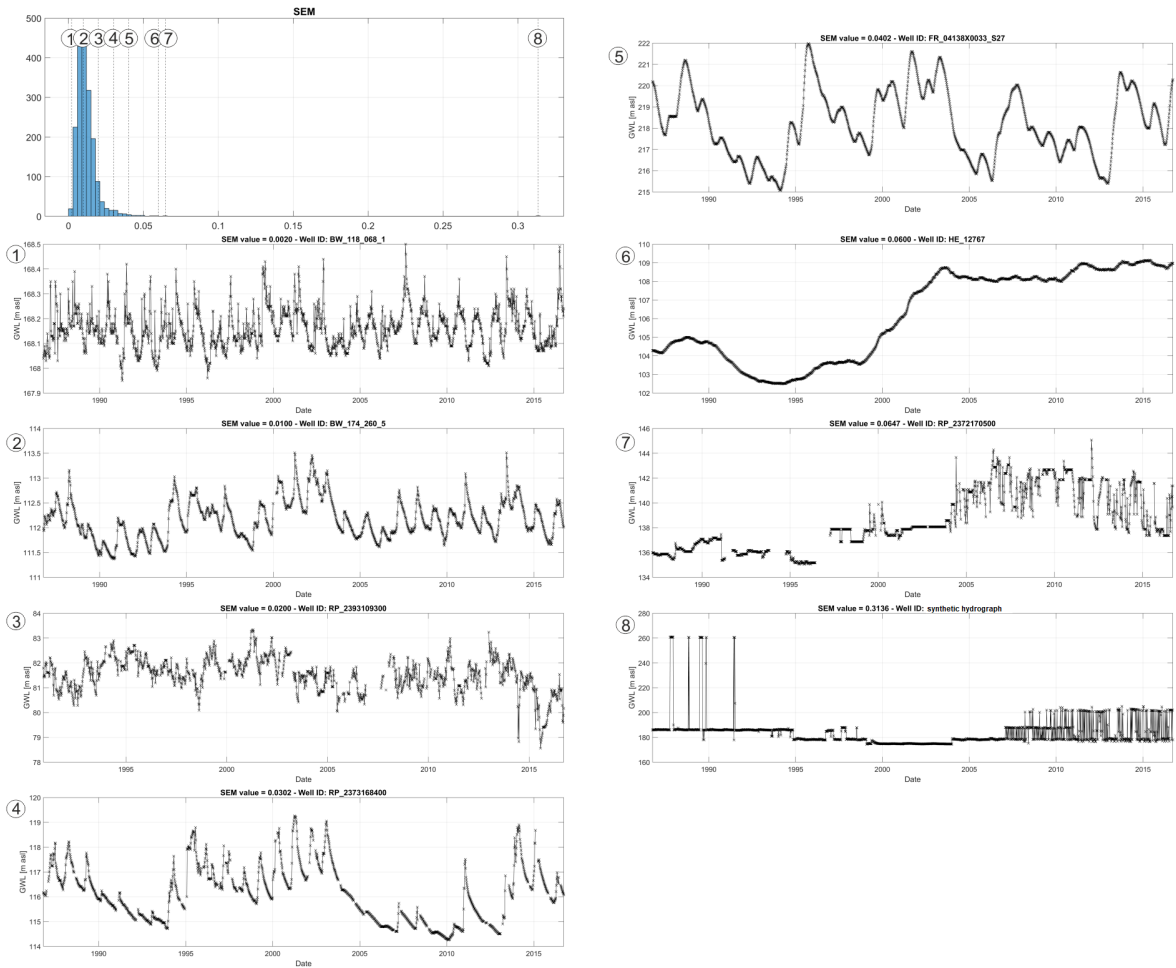
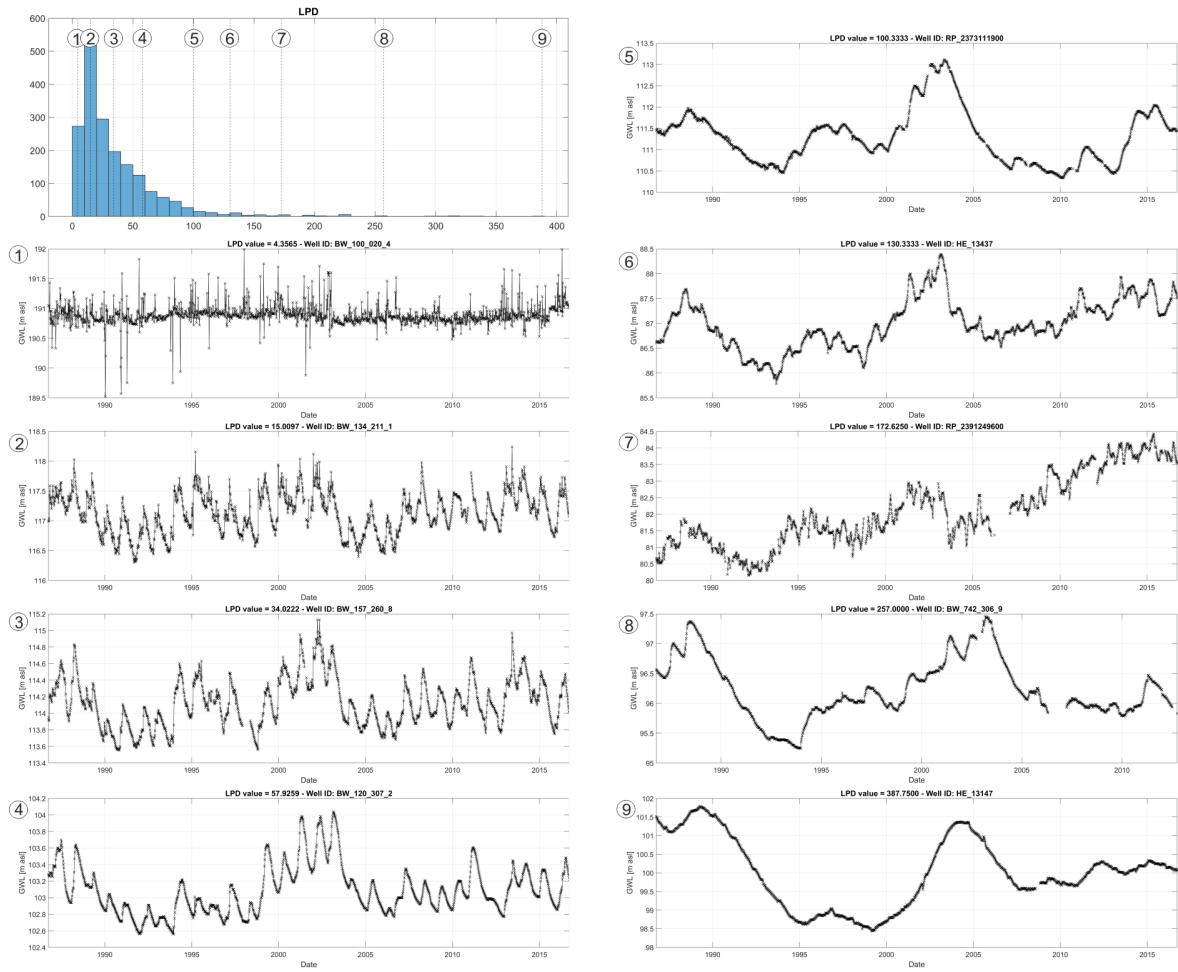


Fig. S11 Selected results of the visual skill test that support the explanatory power of feature  $Y_{var}$ .



**Fig. S12** Selected results of the visual skill test that support the explanatory power of feature SEM.





**Fig. S13** Selected results of the visual skill test that support the explanatory power of feature LPD.

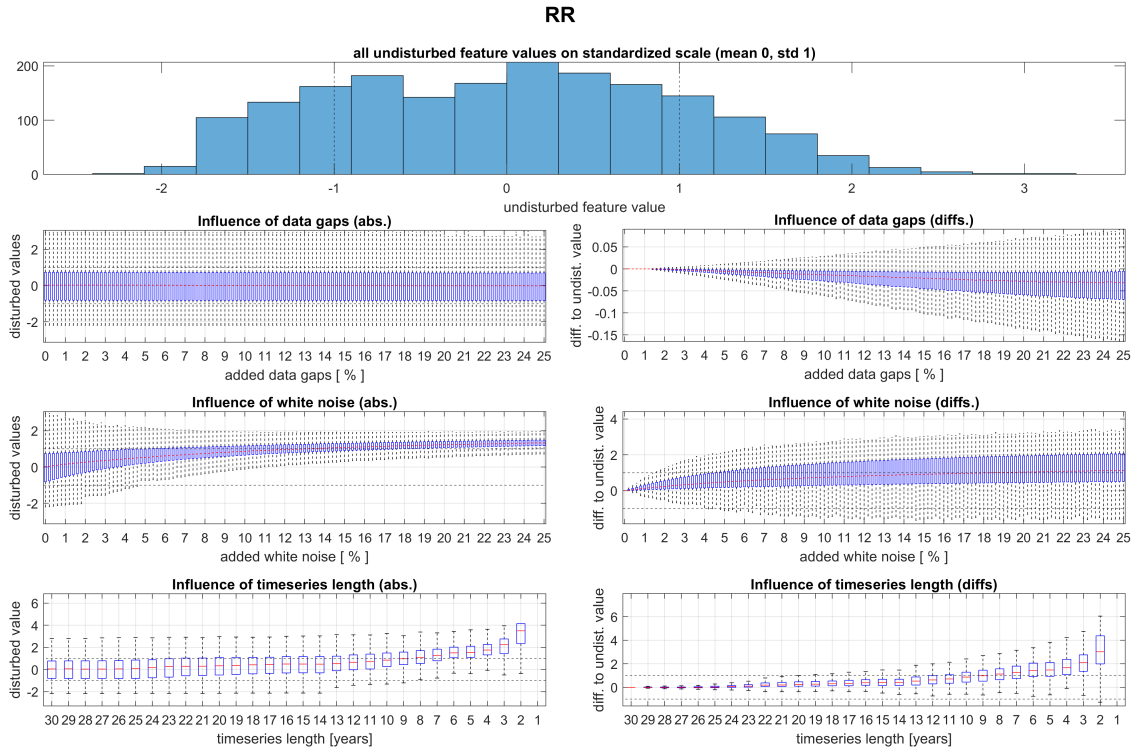


Fig. S14 Results of the three feature robustness experiments for feature RR.

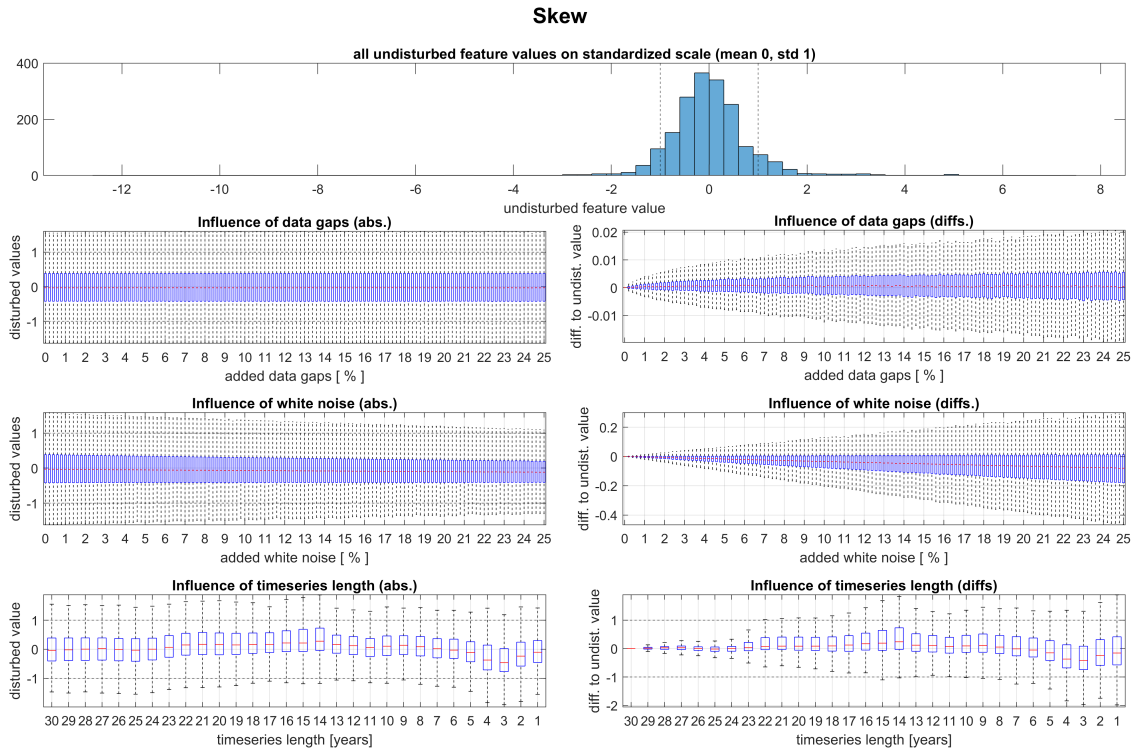


Fig. S15 Results of the three feature robustness experiments for feature Skew.

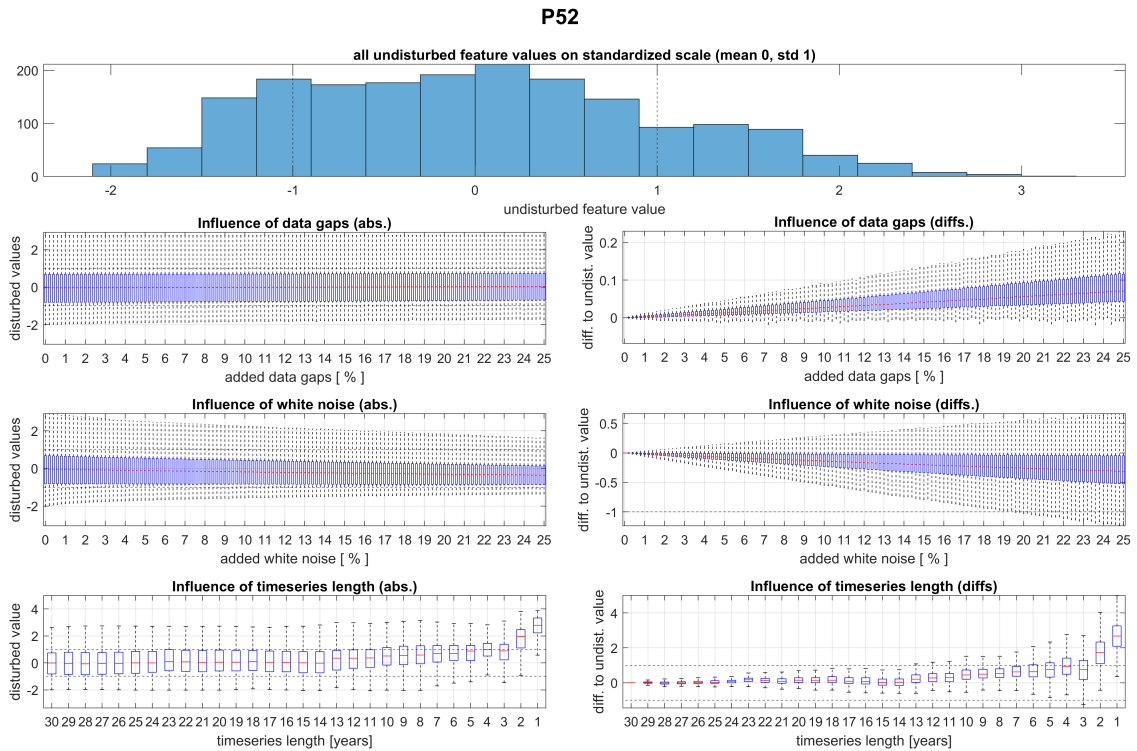


Fig. S16 Results of the three feature robustness experiments for feature P52.

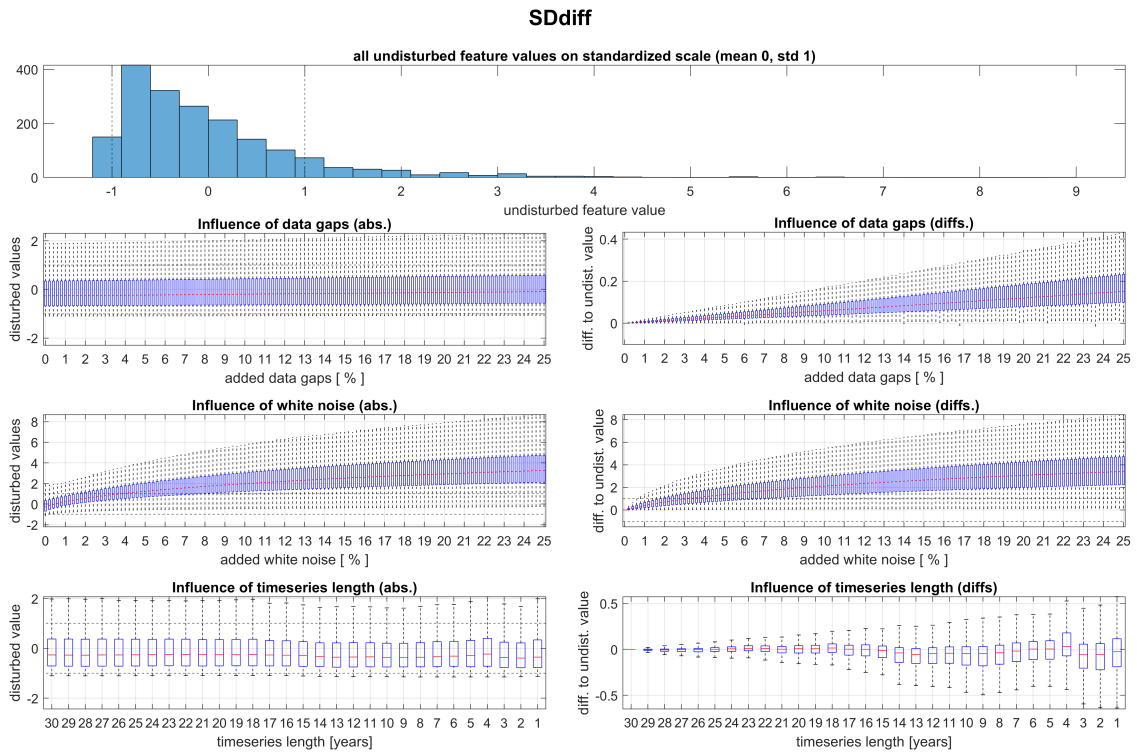
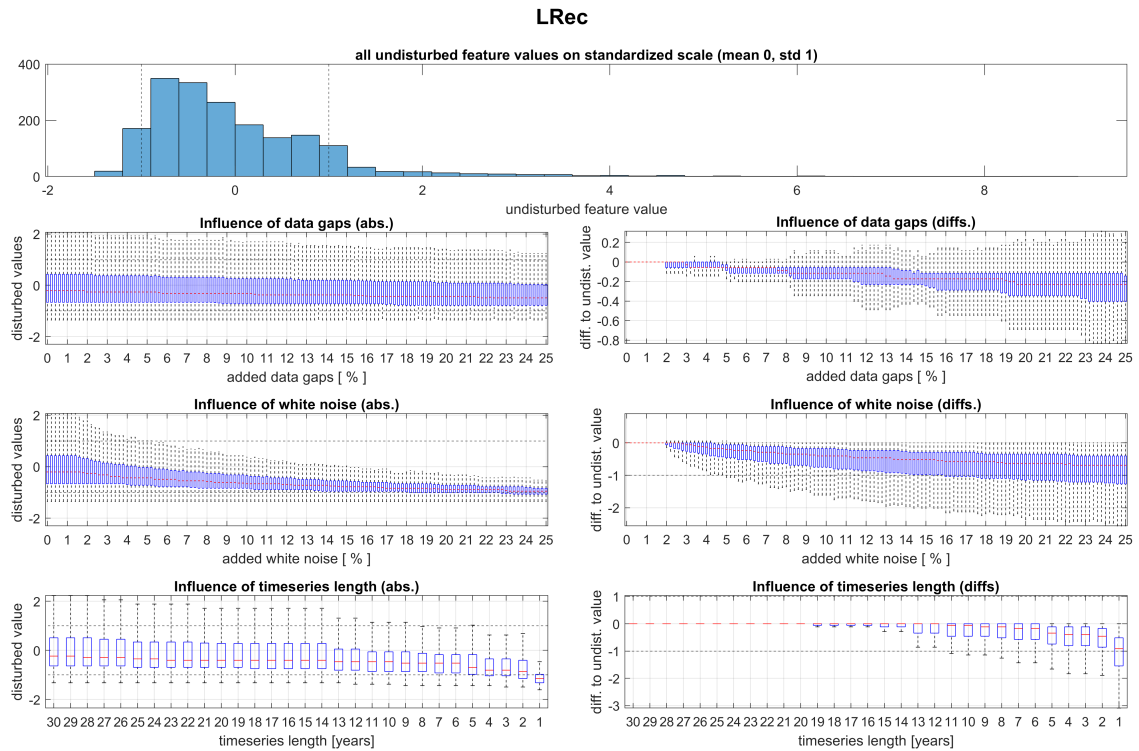
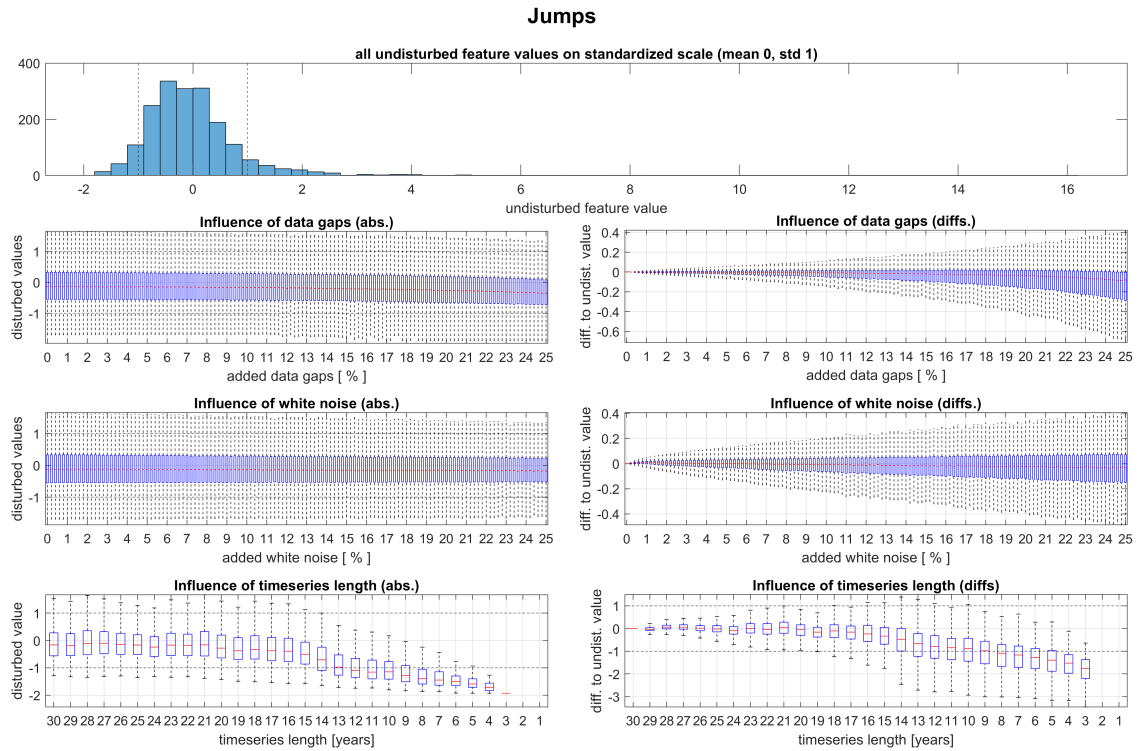


Fig. S17 Results of the three feature robustness experiments for feature  $SD_{diff}$ .



**Fig. S18** Results of the three feature robustness experiments for feature LRec.



**Fig. S19** Results of the three feature robustness experiments for feature Jumps.

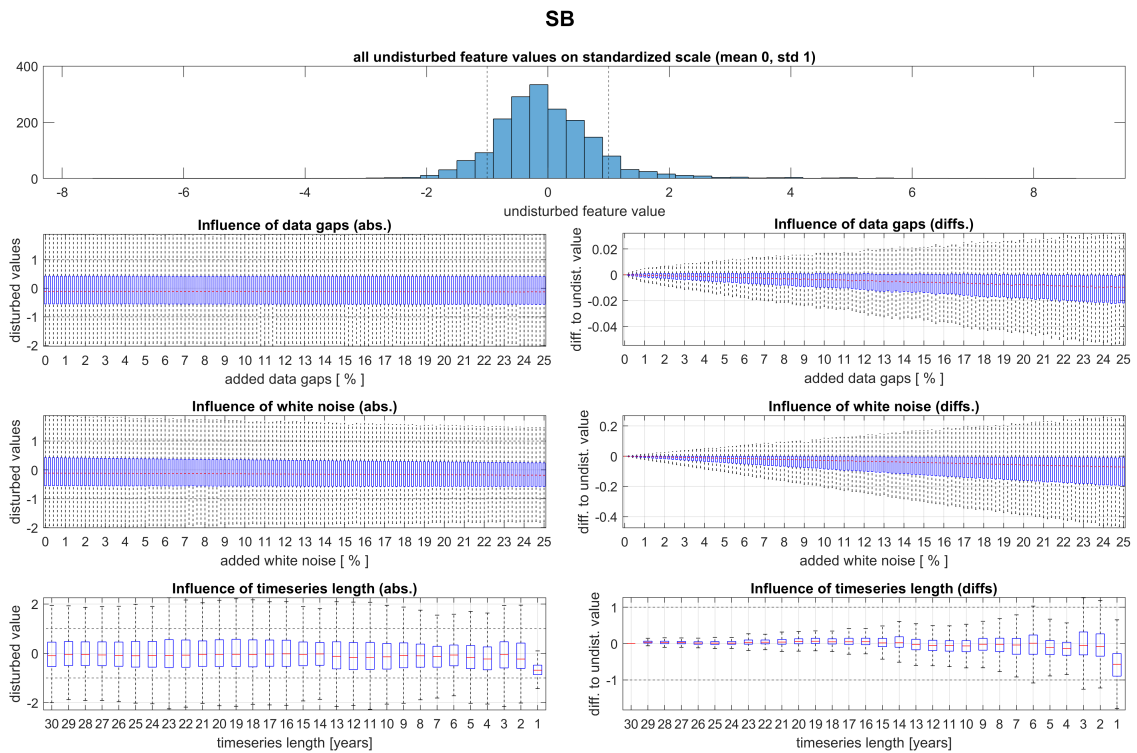


Fig. S20 Results of the three feature robustness experiments for feature SB.

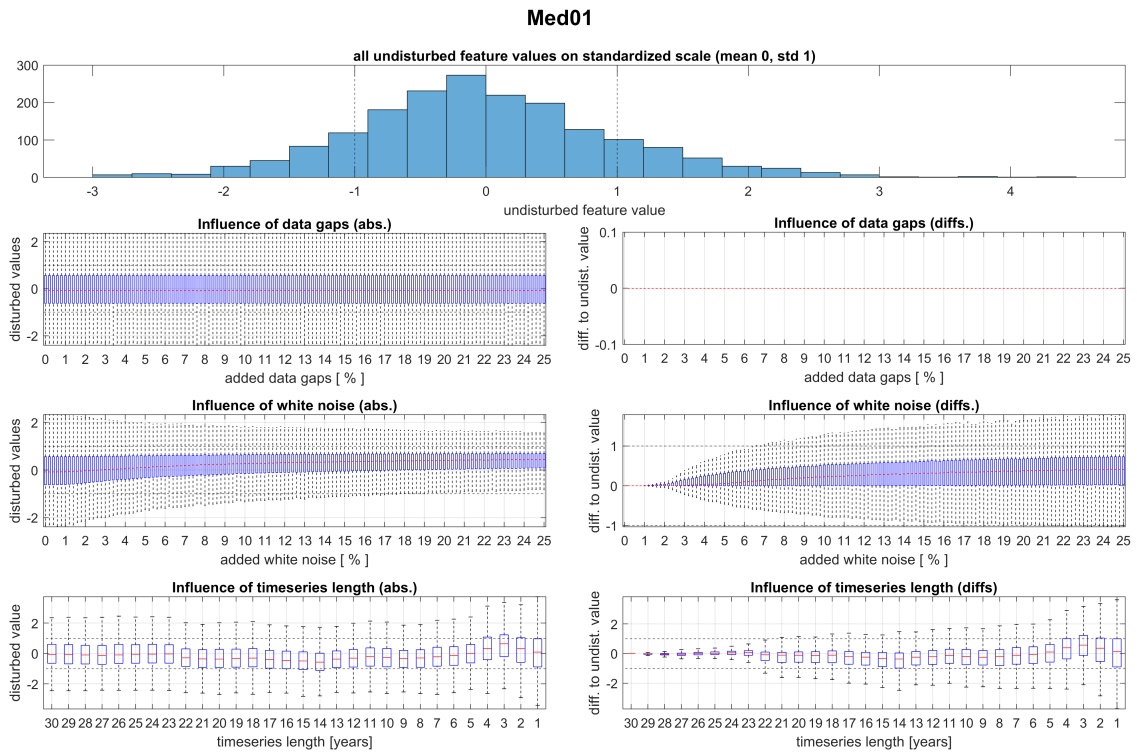


Fig. S21 Results of the three feature robustness experiments for feature Med01.

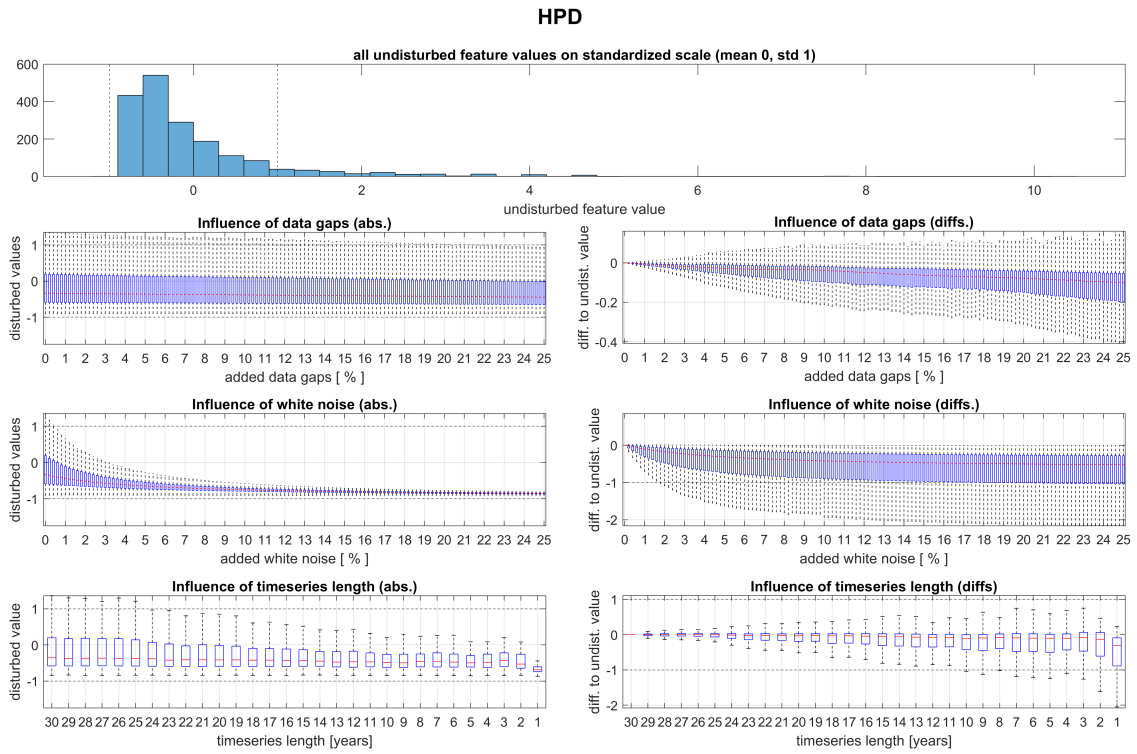


Fig. S22 Results of the three feature robustness experiments for feature HPD.

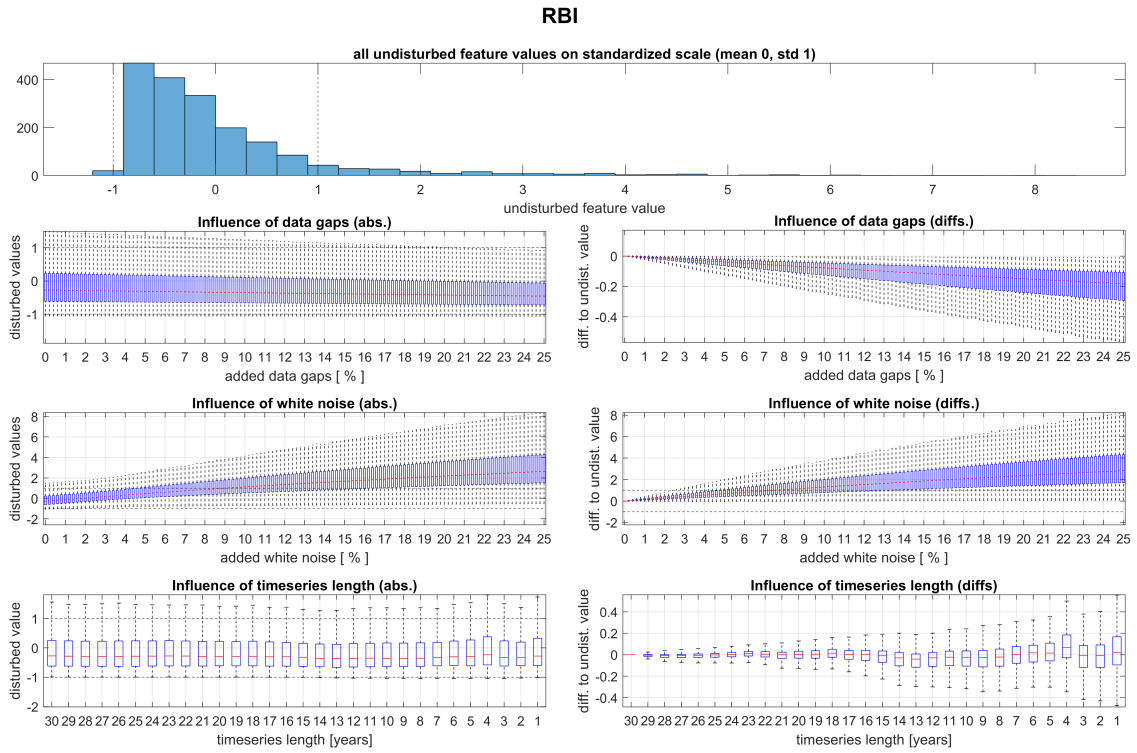


Fig. S23 Results of the three feature robustness experiments for feature RBI.



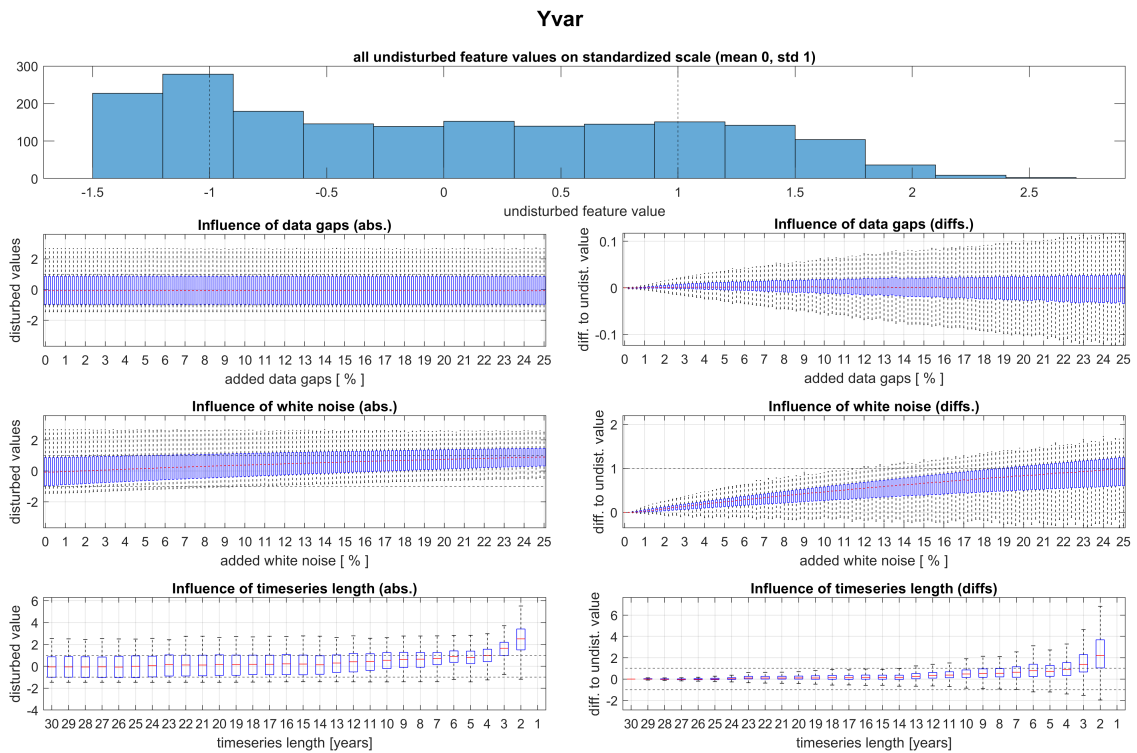


Fig. S24 Results of the three feature robustness experiments for feature  $Y_{var}$ .

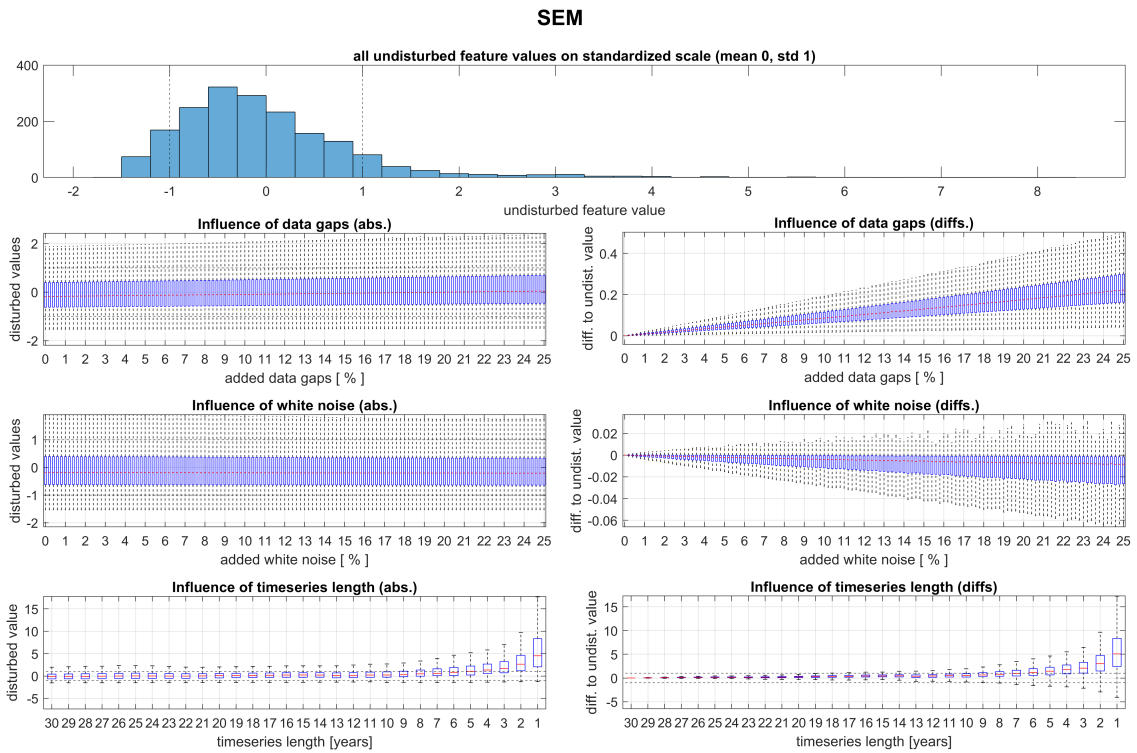
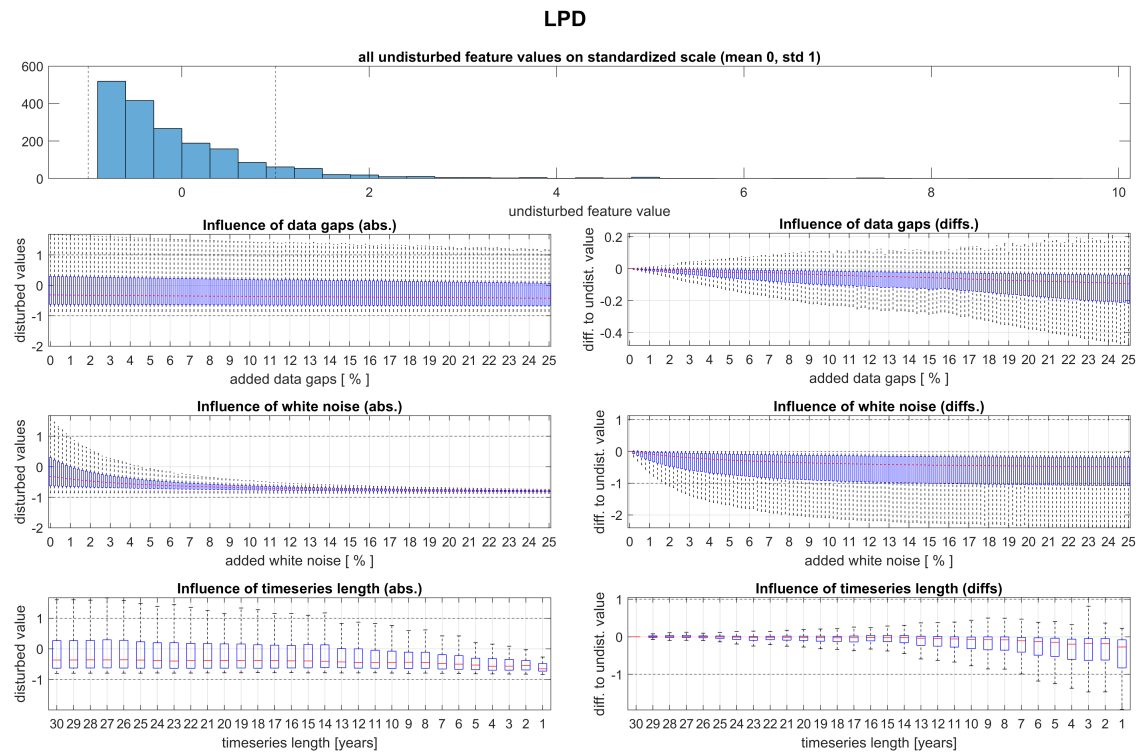


Fig. S25 Results of the three feature robustness experiments for feature SEM.



**Fig. S26** Results of the three feature robustness experiments for feature LPD.



**Table S2** Summary of feature robustness experiments 1 and 2

Feature	Quantile	absolute differences to undisturbed feature value on standardized feature-value scale							
		Added White Noise (%)				Added Data Gaps (%)			
		1	5	10	25	1	5	10	25
<b>RR</b>	2.50%	0.00	0.02	0.03	0.05	0.00	0.00	0.00	0.00
	25%	0.05	0.19	0.30	0.52	0.00	0.00	0.01	0.02
	50%	0.13	0.49	0.75	1.14	0.00	0.01	0.02	0.04
	75%	0.31	1.06	1.51	2.06	0.00	0.01	0.03	0.08
	97.50%	0.57	1.69	2.30	2.92	0.00	0.04	0.08	0.21
<b>Skew</b>	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	25%	0.00	0.01	0.02	0.05	0.00	0.00	0.00	0.00
	50%	0.00	0.03	0.05	0.12	0.00	0.00	0.00	0.01
	75%	0.01	0.04	0.09	0.20	0.00	0.00	0.01	0.01
	97.50%	0.04	0.17	0.32	0.73	0.00	0.02	0.03	0.04
<b>P52</b>	2.50%	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.01
	25%	0.01	0.04	0.08	0.16	0.00	0.01	0.02	0.04
	50%	0.02	0.08	0.15	0.34	0.00	0.01	0.03	0.07
	75%	0.03	0.13	0.24	0.54	0.00	0.02	0.05	0.12
	97.50%	0.07	0.28	0.48	0.97	0.01	0.07	0.13	0.29
<b>SDdiff</b>	2.50%	0.05	0.20	0.36	0.69	0.00	0.01	0.02	0.04
	25%	0.21	0.80	1.31	2.23	0.00	0.02	0.04	0.10
	50%	0.41	1.37	2.10	3.42	0.01	0.03	0.06	0.15
	75%	0.68	1.98	2.95	4.72	0.01	0.05	0.09	0.23
	97.50%	1.57	4.26	6.27	9.68	0.02	0.11	0.22	0.55
<b>Lrec</b>	2.50%	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00
	25%	0.00	0.06	0.17	0.40	0.00	0.03	0.06	0.14
	50%	0.00	0.20	0.40	0.69	0.00	0.06	0.11	0.23
	75%	0.00	0.46	0.75	1.26	0.00	0.09	0.17	0.40
	97.50%	0.35	1.96	2.54	3.29	0.06	0.20	0.40	0.87
<b>Jumps</b>	2.50%	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	25%	0.01	0.01	0.02	0.05	0.00	0.01	0.01	0.04
	50%	0.01	0.03	0.05	0.11	0.00	0.02	0.03	0.12
	75%	0.02	0.06	0.10	0.20	0.01	0.04	0.08	0.30
	97.50%	0.04	0.17	0.29	0.58	0.03	0.26	0.41	1.45
<b>SB</b>	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	25%	0.00	0.01	0.01	0.02	0.00	0.00	0.00	0.00
	50%	0.00	0.02	0.03	0.08	0.00	0.00	0.01	0.01
	75%	0.01	0.05	0.09	0.21	0.00	0.01	0.01	0.02
	97.50%	0.09	0.37	0.69	1.30	0.01	0.04	0.07	0.15
<b>Med01</b>	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	25%	0.00	0.01	0.06	0.20	0.00	0.00	0.00	0.00
	50%	0.00	0.11	0.30	0.49	0.00	0.00	0.00	0.00
	75%	0.01	0.36	0.56	0.78	0.00	0.00	0.00	0.01
	97.50%	0.25	0.93	1.14	1.38	0.01	0.02	0.03	0.04
<b>HPD</b>	2.50%	0.01	0.04	0.06	0.10	0.00	0.00	0.00	0.01
	25%	0.04	0.14	0.21	0.28	0.00	0.01	0.02	0.06
	50%	0.10	0.31	0.42	0.52	0.01	0.03	0.04	0.10
	75%	0.31	0.74	0.90	1.04	0.01	0.06	0.10	0.20
	97.50%	2.03	3.22	3.49	3.69	0.05	0.24	0.42	0.86
<b>RBI</b>	2.50%	0.00	0.02	0.05	0.15	0.00	0.00	0.00	0.00
	25%	0.01	0.13	0.27	0.60	0.00	0.00	0.00	0.01
	50%	0.03	0.24	0.46	0.99	0.00	0.01	0.01	0.03
	75%	0.06	0.31	0.58	1.24	0.00	0.02	0.04	0.08
	97.50%	0.13	0.42	0.73	1.48	0.01	0.09	0.15	0.32
<b>Yvar</b>	2.50%	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.08
	25%	0.00	0.00	0.00	0.00	0.01	0.03	0.06	0.16
	50%	0.00	0.00	0.00	0.01	0.01	0.04	0.09	0.22
	75%	0.00	0.01	0.01	0.03	0.01	0.06	0.11	0.30
	97.50%	0.02	0.09	0.18	0.36	0.02	0.12	0.23	0.61
<b>SEM</b>	2.50%	0.00	0.02	0.03	0.06	0.00	0.00	0.00	0.01
	25%	0.02	0.08	0.13	0.20	0.00	0.01	0.02	0.04
	50%	0.08	0.27	0.38	0.48	0.01	0.03	0.04	0.10
	75%	0.29	0.73	0.92	1.08	0.01	0.06	0.10	0.22
	97.50%	1.55	2.75	3.07	3.30	0.05	0.20	0.37	0.77
<b>LPD</b>	2.50%	0.02	0.12	0.22	0.49	0.00	0.01	0.02	0.05
	25%	0.09	0.41	0.79	1.71	0.00	0.02	0.05	0.11
	50%	0.14	0.68	1.31	2.82	0.01	0.04	0.08	0.18
	75%	0.22	1.05	2.00	4.32	0.01	0.06	0.13	0.29
	97.50%	0.40	1.91	3.64	7.87	0.04	0.20	0.39	0.91

Table S3 Summary of feature robustness experiment 3

Feature	Quantile	absolute differences to undisturbed feature value on standardized feature-value scale							
		Timeseries length [years]							
		30	25	20	15	10	5	3	1
RR	2.50%	0.00	0.00	0.01	0.01	0.06	0.22	0.23	NaN
	25%	0.00	0.03	0.11	0.18	0.45	0.93	1.36	NaN
	50%	0.00	0.05	0.24	0.40	0.87	1.47	2.10	NaN
	75%	0.00	0.11	0.47	0.69	1.28	2.11	2.75	NaN
	97.50%	0.00	0.50	1.00	1.31	1.93	3.25	4.09	NaN
Skew	2.50%	0.00	0.00	0.01	0.01	0.01	0.01	0.03	0.03
	25%	0.00	0.04	0.07	0.12	0.11	0.19	0.32	0.26
	50%	0.00	0.08	0.17	0.29	0.27	0.39	0.53	0.51
	75%	0.00	0.14	0.44	0.65	0.59	0.61	0.82	0.81
	97.50%	0.00	0.59	1.26	1.74	1.67	1.90	1.89	2.31
P52	2.50%	0.00	0.00	0.01	0.01	0.02	0.04	0.05	0.47
	25%	0.00	0.03	0.08	0.08	0.24	0.36	0.36	2.09
	50%	0.00	0.07	0.16	0.18	0.45	0.78	0.78	2.68
	75%	0.00	0.13	0.29	0.31	0.73	1.14	1.28	3.26
	97.50%	0.00	0.30	0.66	0.78	1.50	1.89	2.37	4.87
SDdiff	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
	25%	0.00	0.01	0.02	0.02	0.04	0.04	0.06	0.07
	50%	0.00	0.02	0.04	0.06	0.09	0.10	0.14	0.16
	75%	0.00	0.05	0.09	0.15	0.22	0.28	0.31	0.37
	97.50%	0.00	0.25	0.45	0.63	0.93	1.07	1.21	1.27
Lrec	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11
	25%	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.52
	50%	0.00	0.00	0.00	0.00	0.06	0.34	0.40	0.92
	75%	0.00	0.00	0.00	0.11	0.46	0.74	0.80	1.55
	97.50%	0.00	0.75	1.21	1.37	2.29	2.41	2.69	3.56
Jumps	2.50%	0.00	0.00	0.01	0.02	0.05	0.29	0.89	NaN
	25%	0.00	0.07	0.11	0.23	0.45	0.96	1.37	NaN
	50%	0.00	0.14	0.23	0.44	0.91	1.39	1.76	NaN
	75%	0.00	0.25	0.43	0.76	1.44	1.88	2.21	NaN
	97.50%	0.00	0.77	1.49	2.29	3.31	3.58	4.10	NaN
SB	2.50%	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03
	25%	0.00	0.02	0.03	0.06	0.07	0.10	0.16	0.38
	50%	0.00	0.04	0.07	0.12	0.16	0.22	0.32	0.61
	75%	0.00	0.07	0.15	0.21	0.28	0.43	0.53	0.92
	97.50%	0.00	0.26	0.69	0.69	0.78	1.19	1.36	2.52
Med01	2.50%	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.04
	25%	0.00	0.04	0.11	0.20	0.24	0.27	0.45	0.48
	50%	0.00	0.10	0.29	0.46	0.53	0.60	0.85	0.95
	75%	0.00	0.19	0.66	0.93	0.94	1.05	1.38	1.55
	97.50%	0.00	0.68	1.66	2.12	2.14	2.24	2.57	3.14
HPD	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
	25%	0.00	0.02	0.03	0.03	0.05	0.05	0.07	0.10
	50%	0.00	0.05	0.08	0.10	0.14	0.15	0.17	0.31
	75%	0.00	0.15	0.27	0.33	0.45	0.51	0.51	0.89
	97.50%	0.00	1.57	2.57	2.41	3.13	2.97	3.15	3.71
RBI	2.50%	0.00	0.00	0.01	0.01	0.02	0.02	0.09	NaN
	25%	0.00	0.03	0.06	0.09	0.26	0.33	0.71	NaN
	50%	0.00	0.06	0.15	0.21	0.51	0.76	1.38	NaN
	75%	0.00	0.12	0.32	0.40	0.88	1.29	2.29	NaN
	97.50%	0.00	0.45	0.89	0.98	1.65	2.53	3.97	NaN
Yvar	2.50%	0.00	0.02	0.03	0.06	0.06	0.11	0.13	0.20
	25%	0.00	0.09	0.14	0.30	0.34	0.76	1.14	2.45
	50%	0.00	0.15	0.26	0.47	0.60	1.49	2.13	5.08
	75%	0.00	0.23	0.40	0.70	1.00	2.30	3.33	8.38
	97.50%	0.00	0.49	0.88	1.54	2.32	5.39	6.90	19.77
SEM	2.50%	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	25%	0.00	0.01	0.02	0.03	0.04	0.06	0.07	0.09
	50%	0.00	0.03	0.05	0.08	0.12	0.17	0.19	0.27
	75%	0.00	0.13	0.17	0.25	0.35	0.53	0.64	0.83
	97.50%	0.00	1.35	1.78	2.35	2.36	2.82	3.09	3.10
LPD	2.50%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	25%	0.00	0.01	0.02	0.02	0.03	0.04	0.05	0.05
	50%	0.00	0.02	0.03	0.05	0.07	0.09	0.10	0.13
	75%	0.00	0.04	0.07	0.11	0.16	0.19	0.21	0.26
	97.50%	0.00	0.15	0.27	0.48	0.69	0.70	0.70	0.95

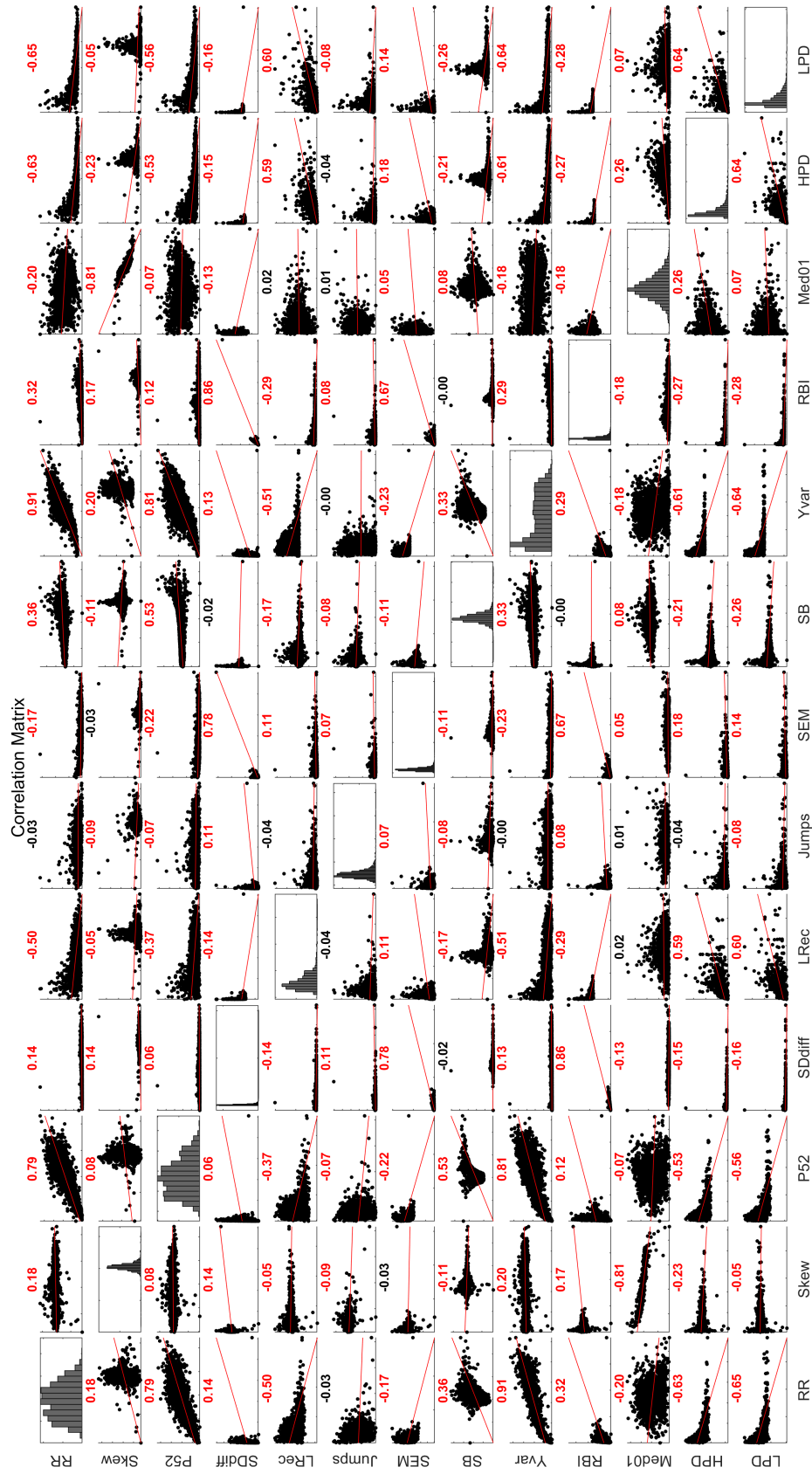
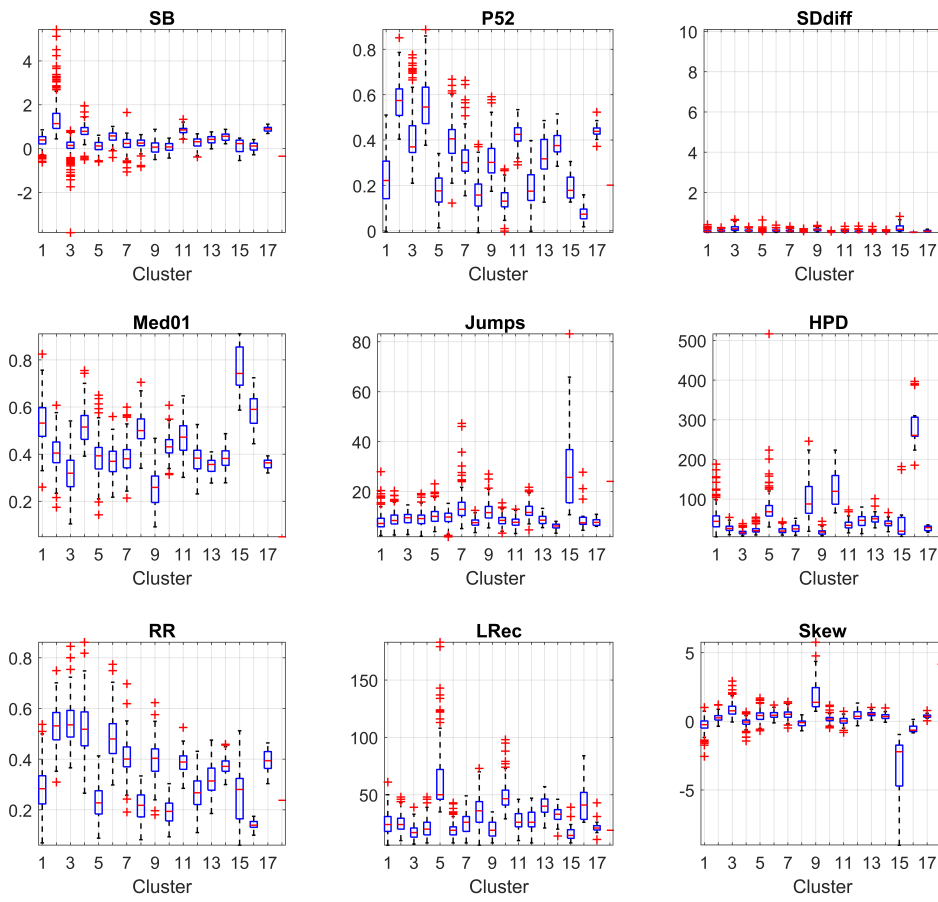


Fig. S27 Detailed Correlation Matrix of all Features, red = significant



**Fig. S28** Boxplots showing feature-value distributions within all clusters including Cluster 18

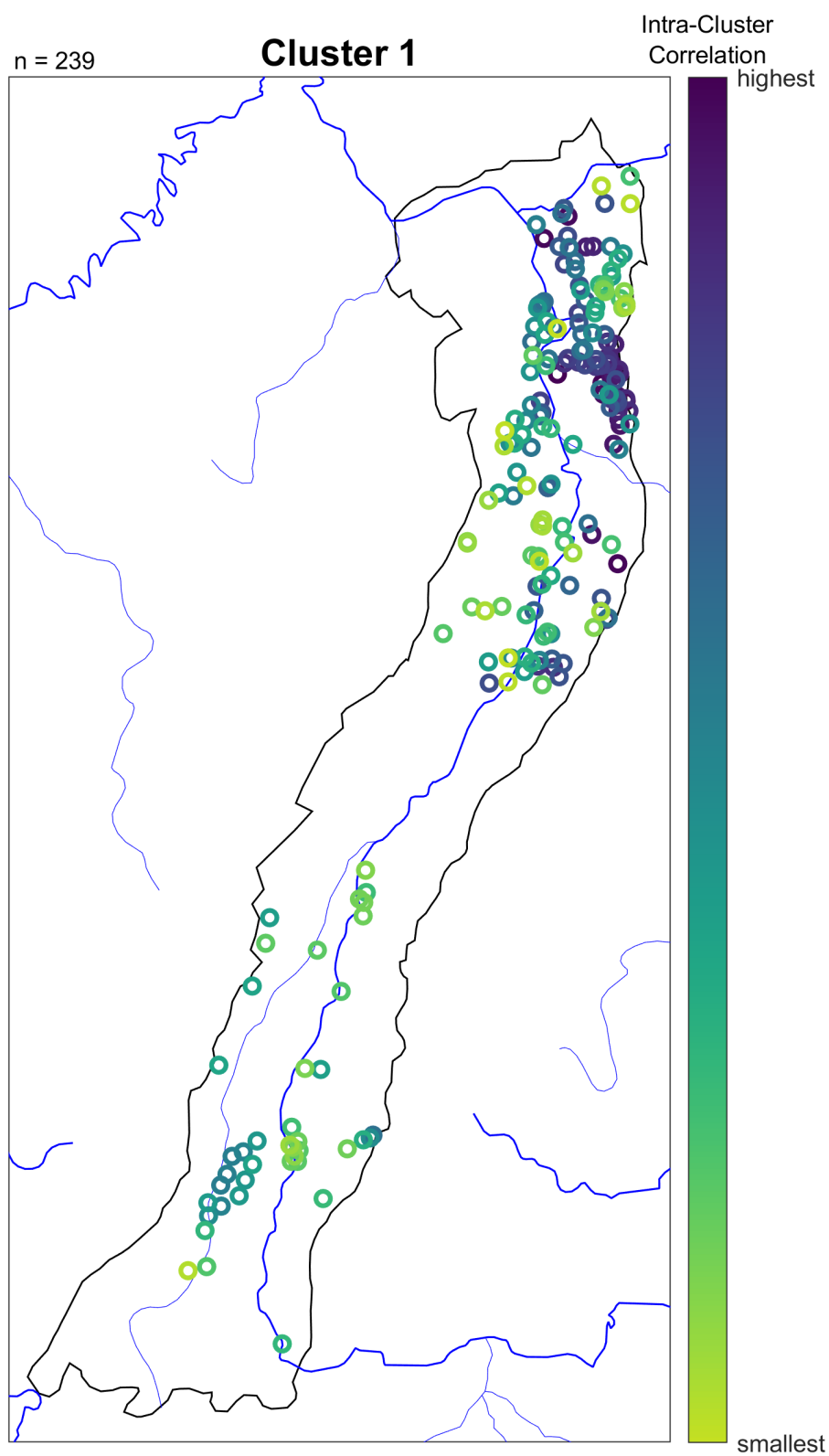


Fig. S29 Well locations in Cluster 1

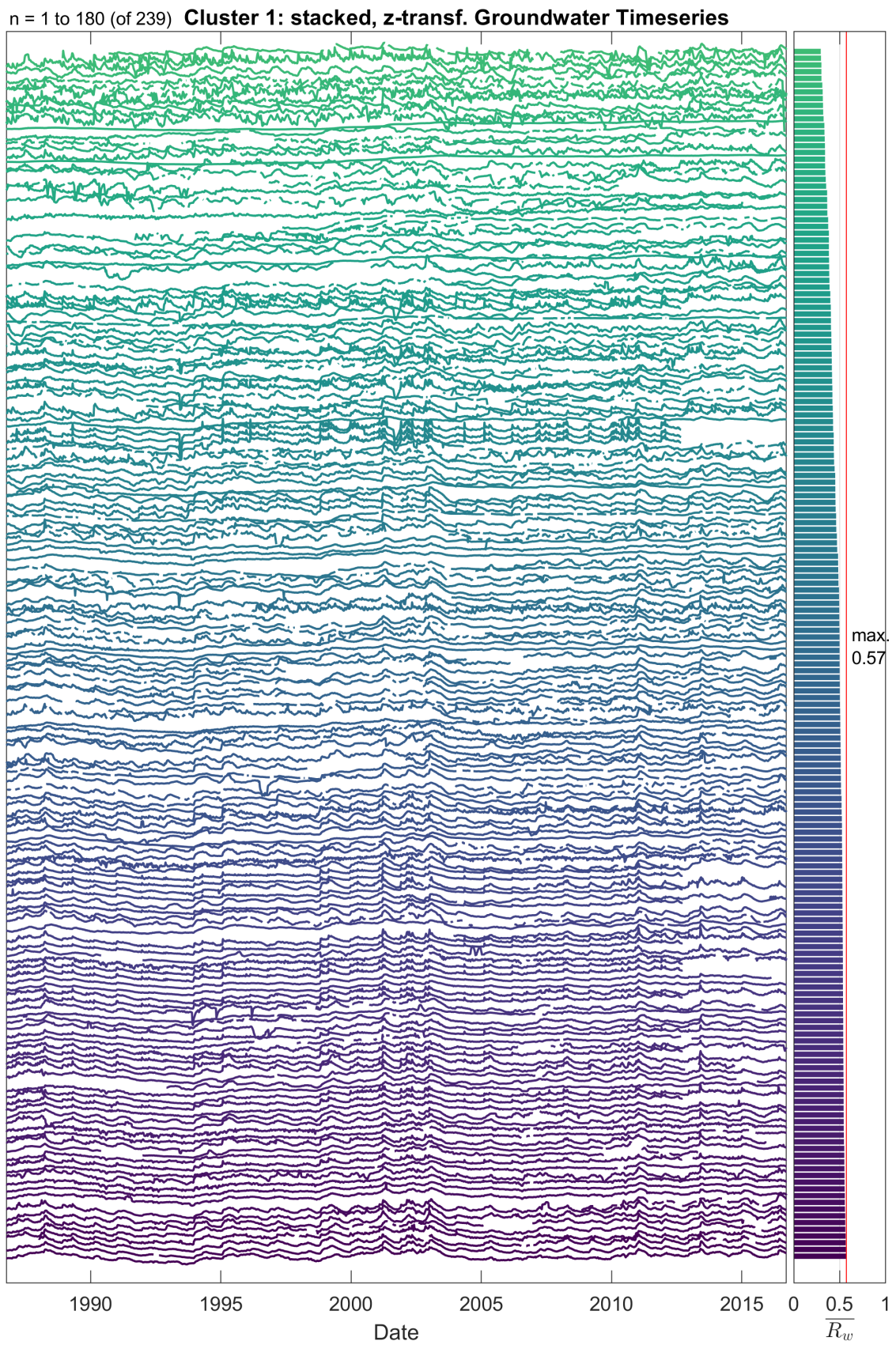
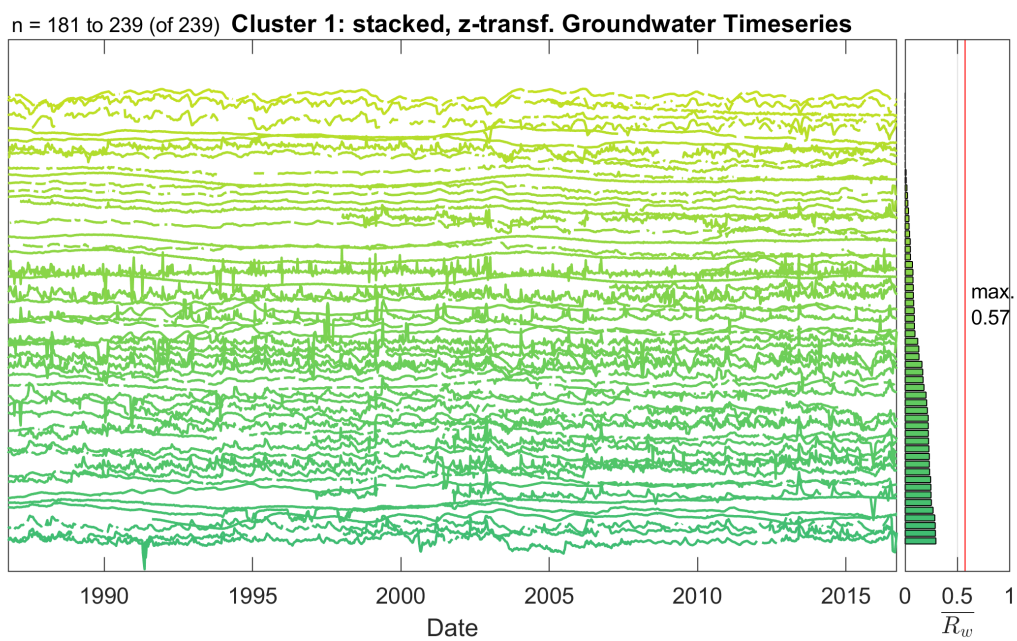
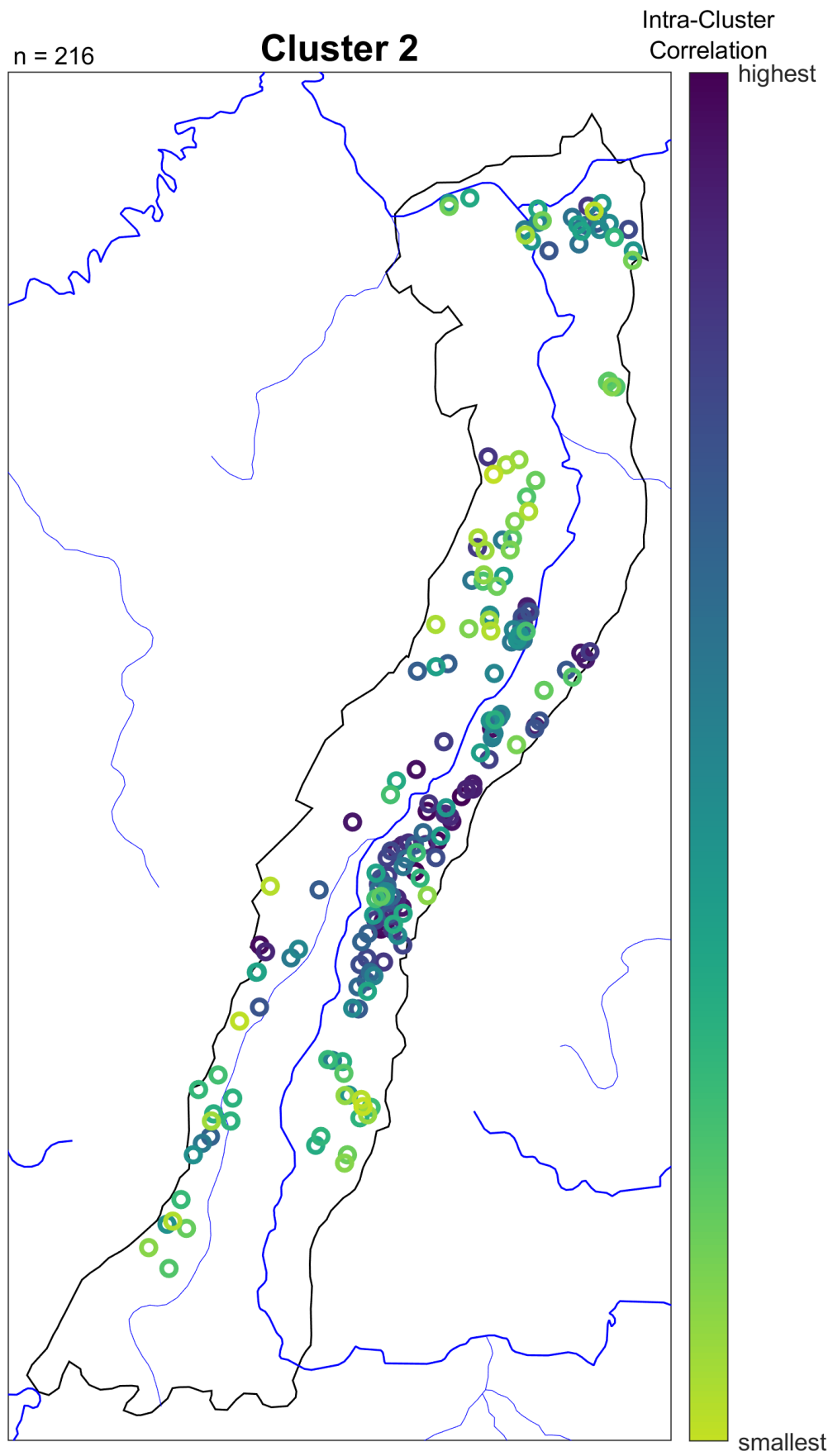


Fig. S30 First 180 stacked, and z-scored hydrographs of Cluster 1



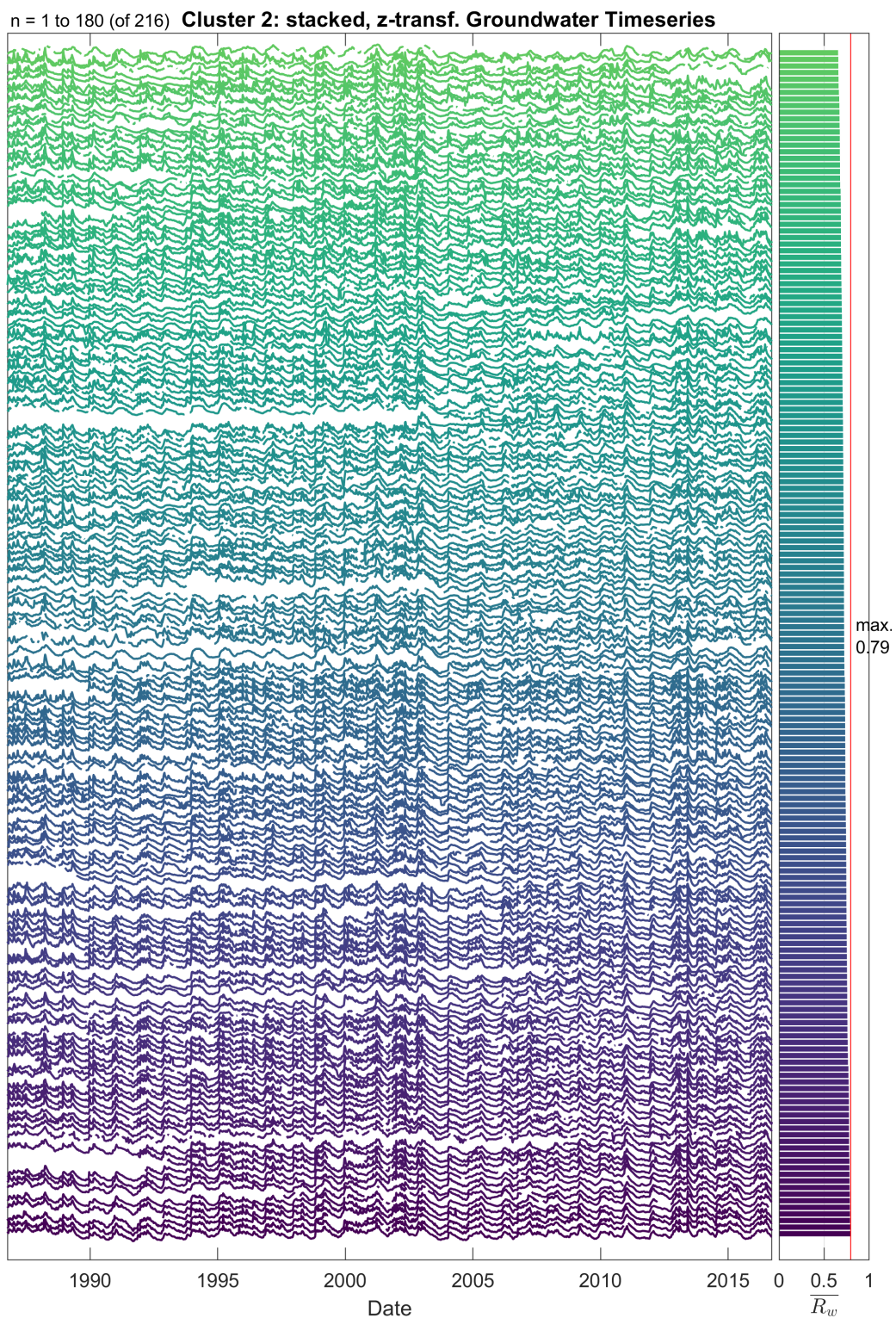


**Fig. S31** Last 59 stacked, and z-scored hydrographs of Cluster 1

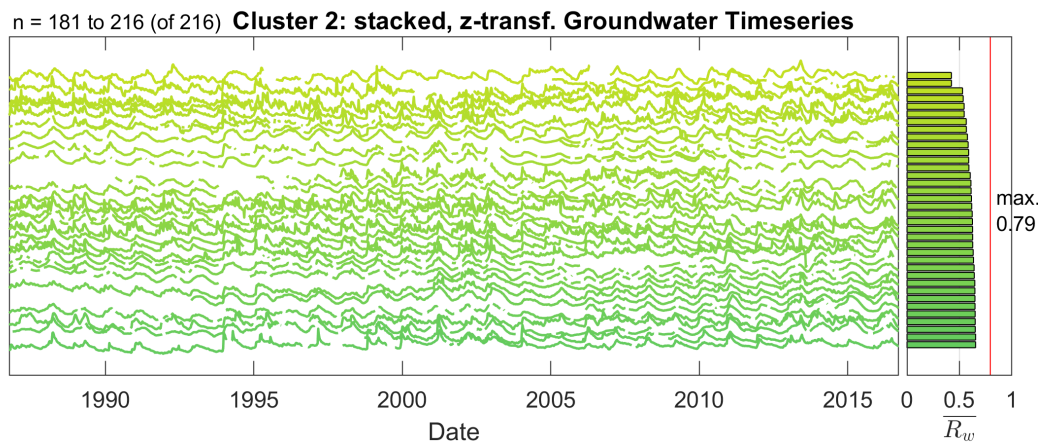


**Fig. S32** Well locations in Cluster 2





**Fig. S33** First 180 stacked, and z-scored hydrographs of Cluster 2



**Fig. S34** Last 36 stacked, and z-scored hydrographs of Cluster 2

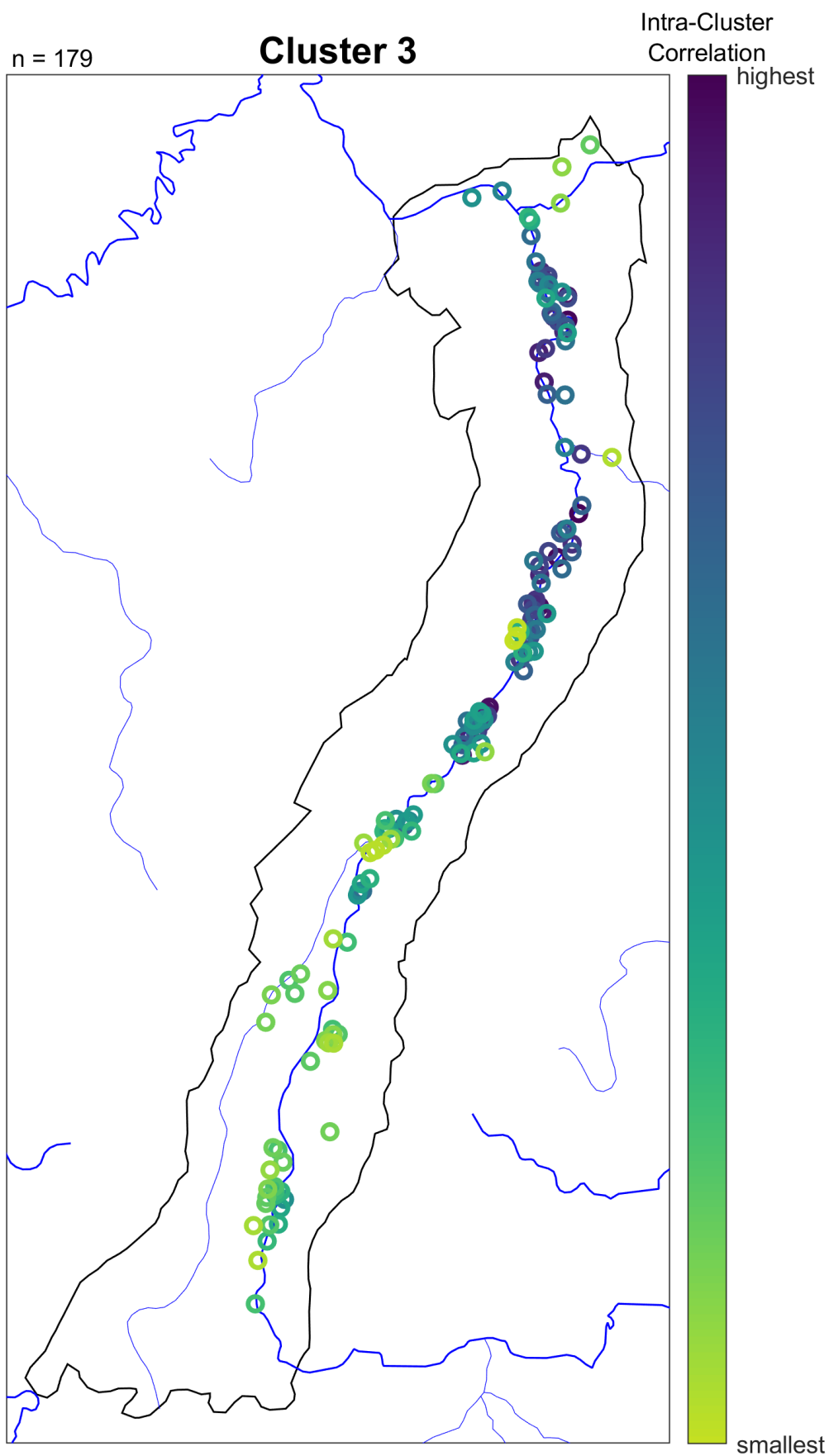
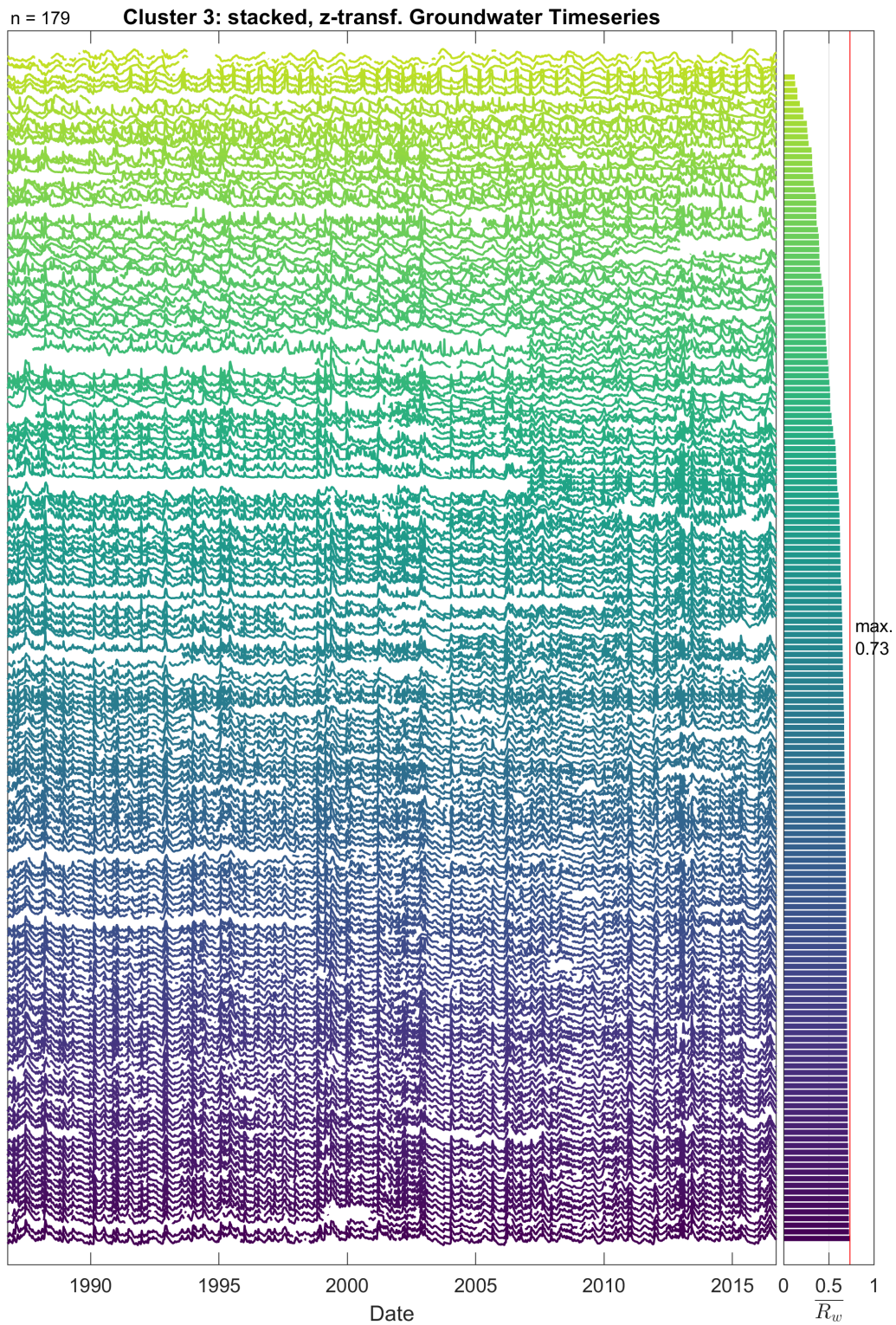


Fig. S35 Well locations in Cluster 3





**Fig. S36** Stacked, and z-scored hydrographs of Cluster 3

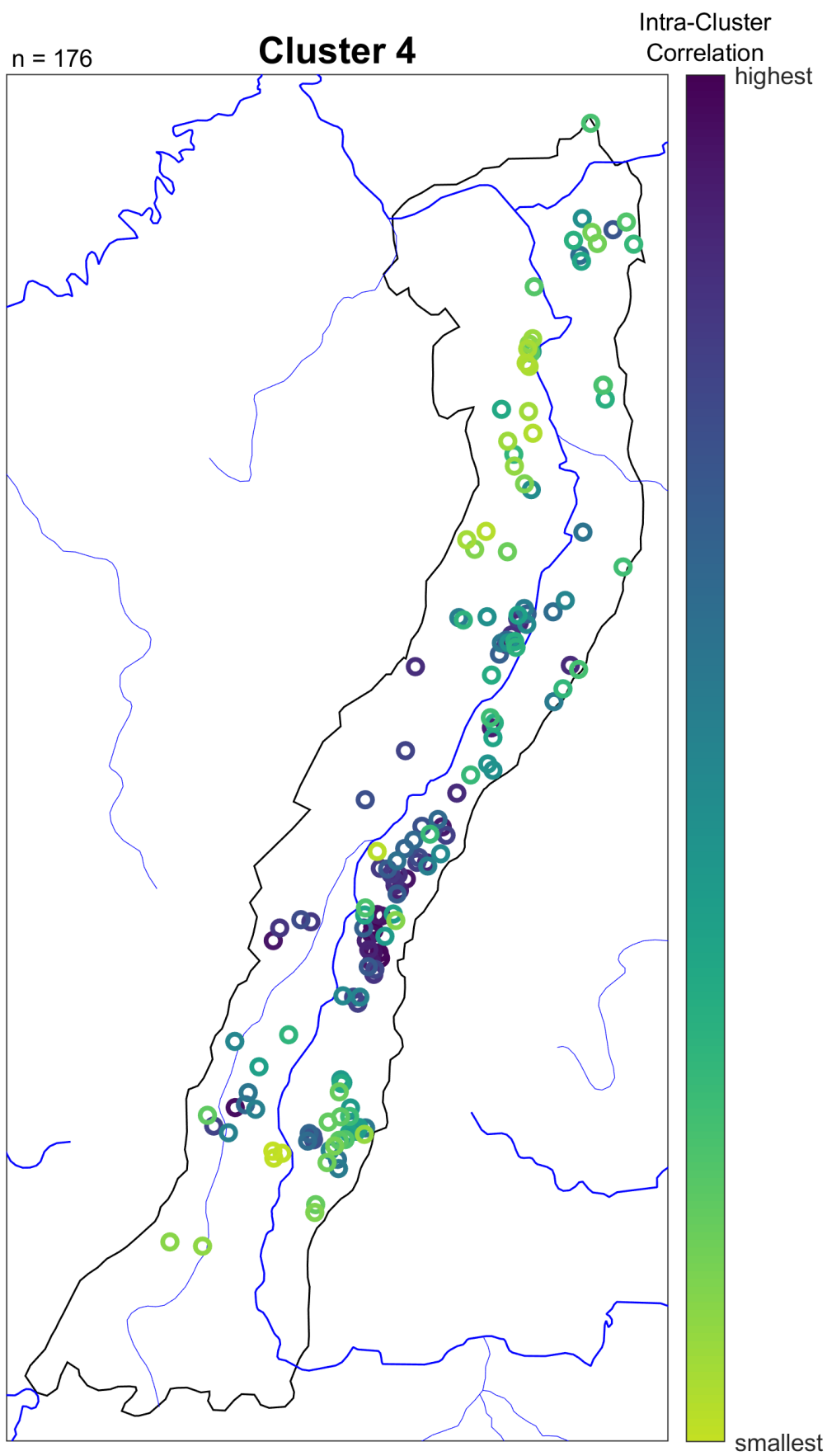
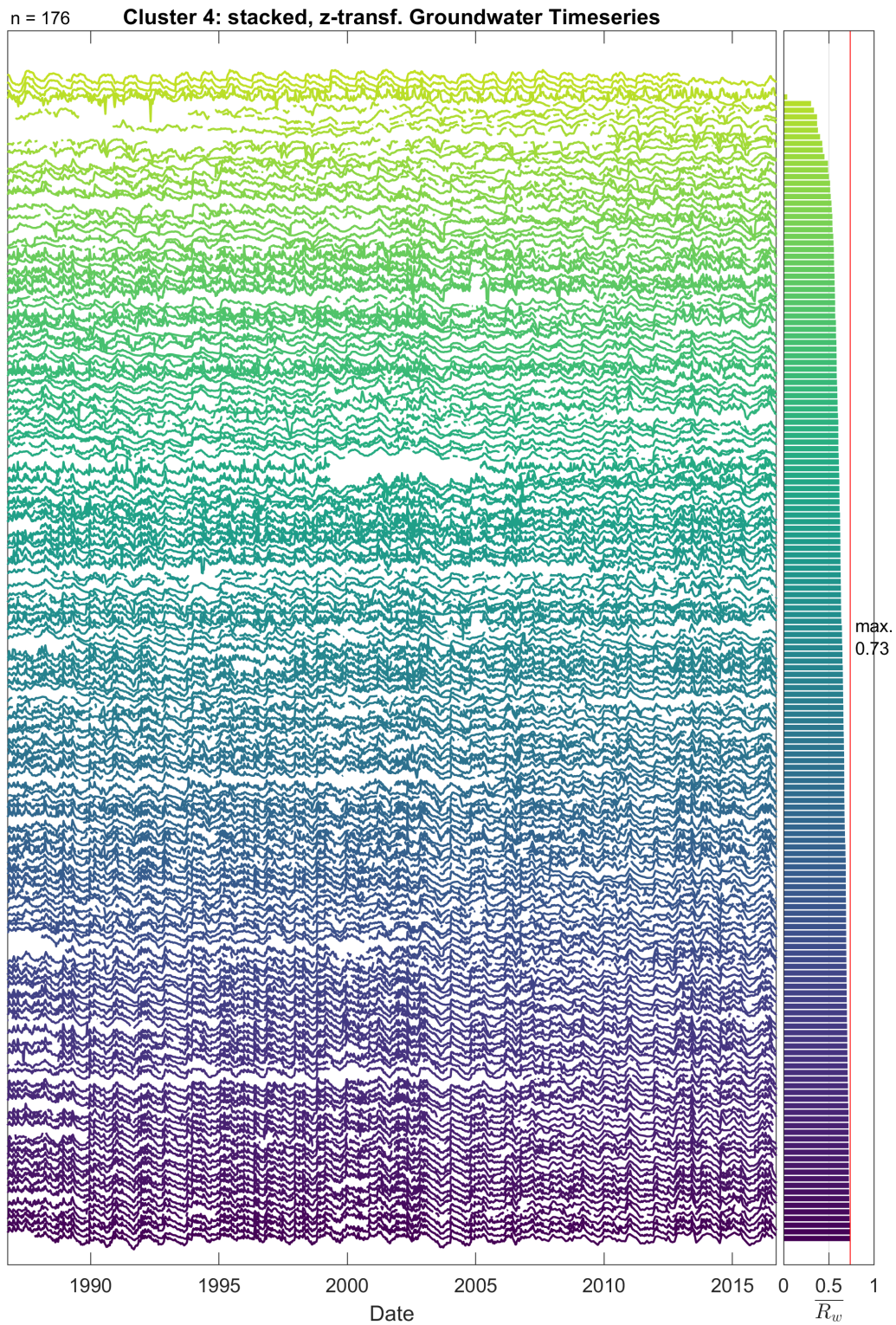


Fig. S37 Well locations in Cluster 4





**Fig. S38** Stacked, and z-scored hydrographs of Cluster 4

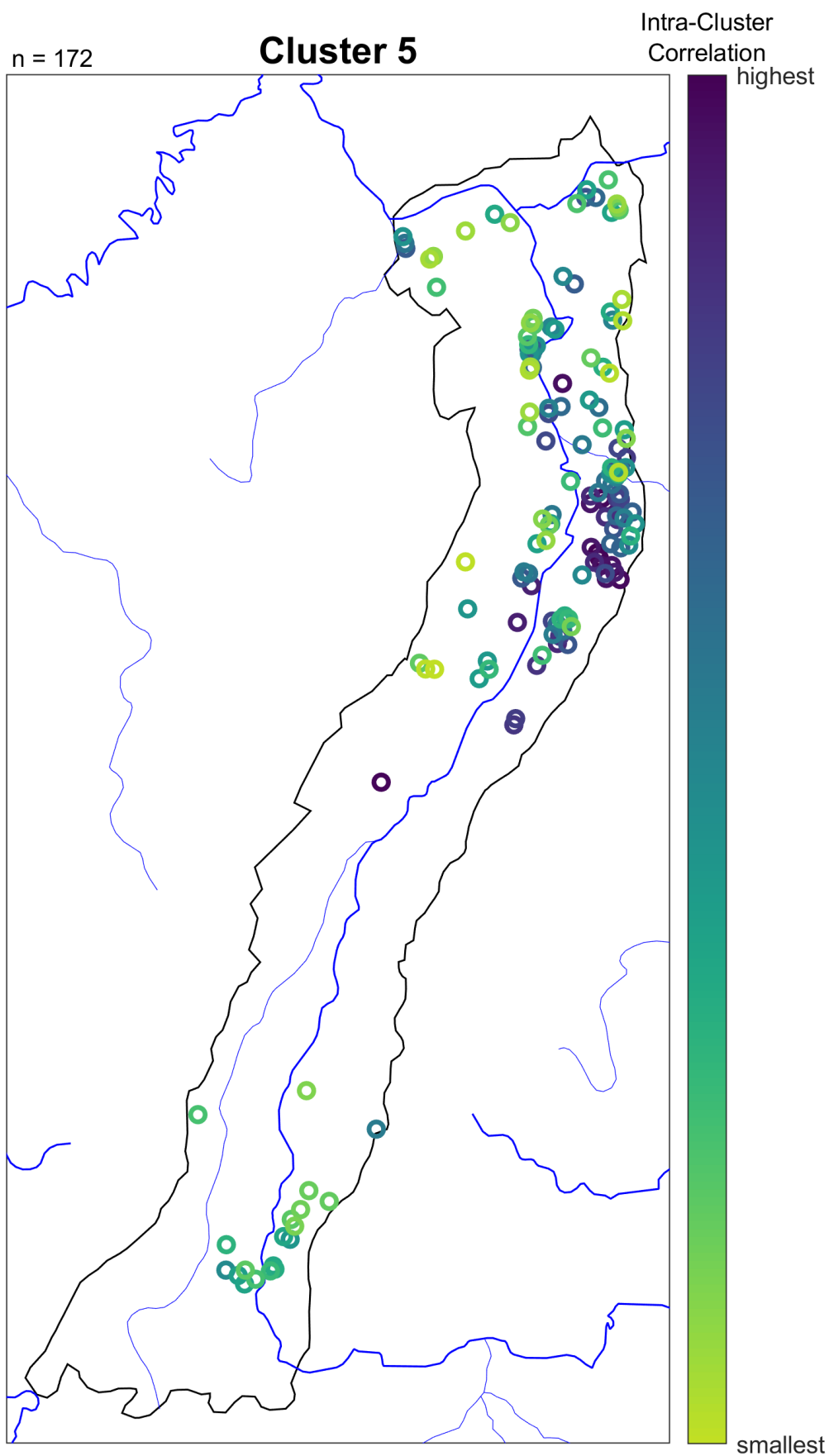
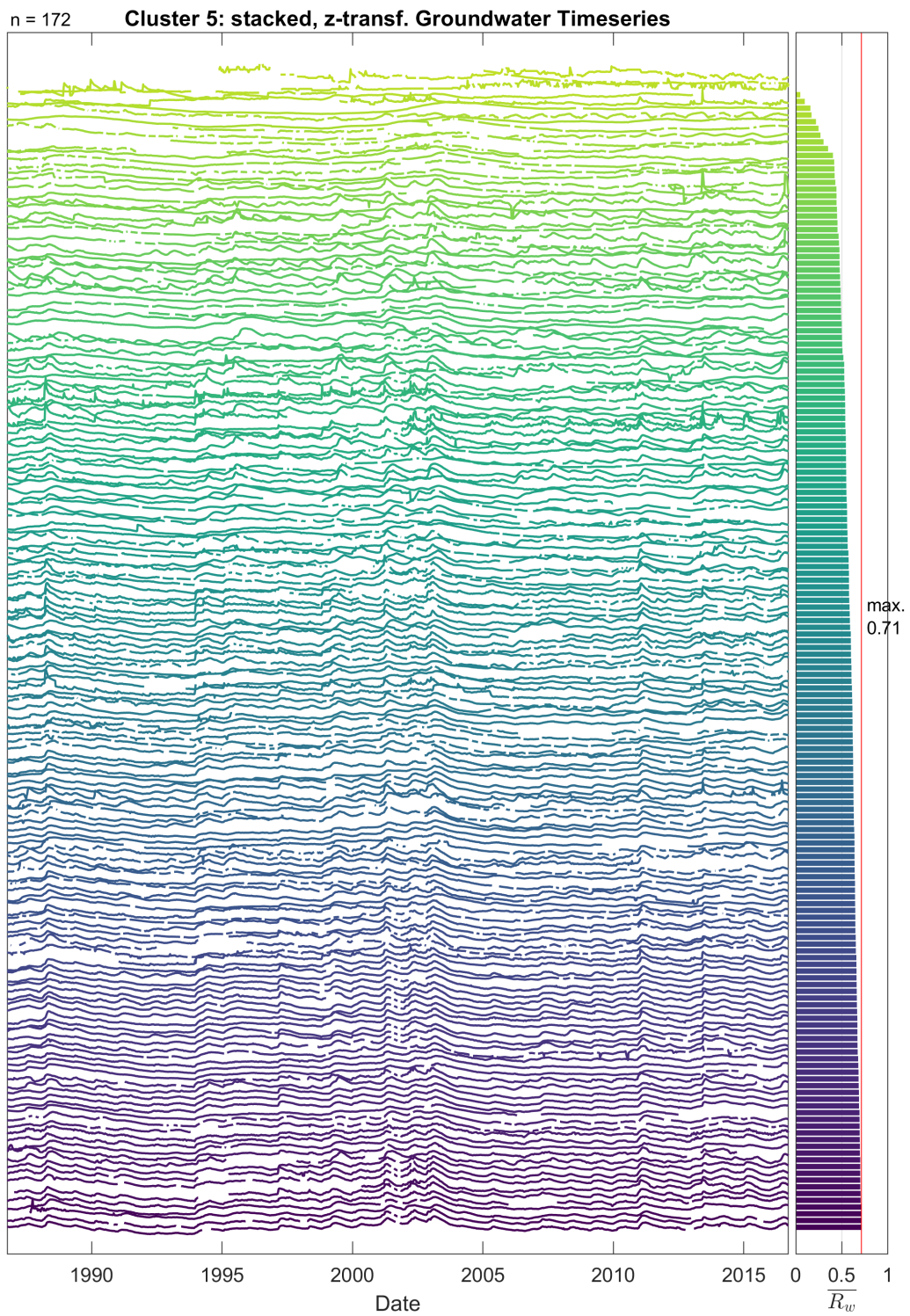


Fig. S39 Well locations in Cluster 5





**Fig. S40** Stacked, and z-scored hydrographs of Cluster 5



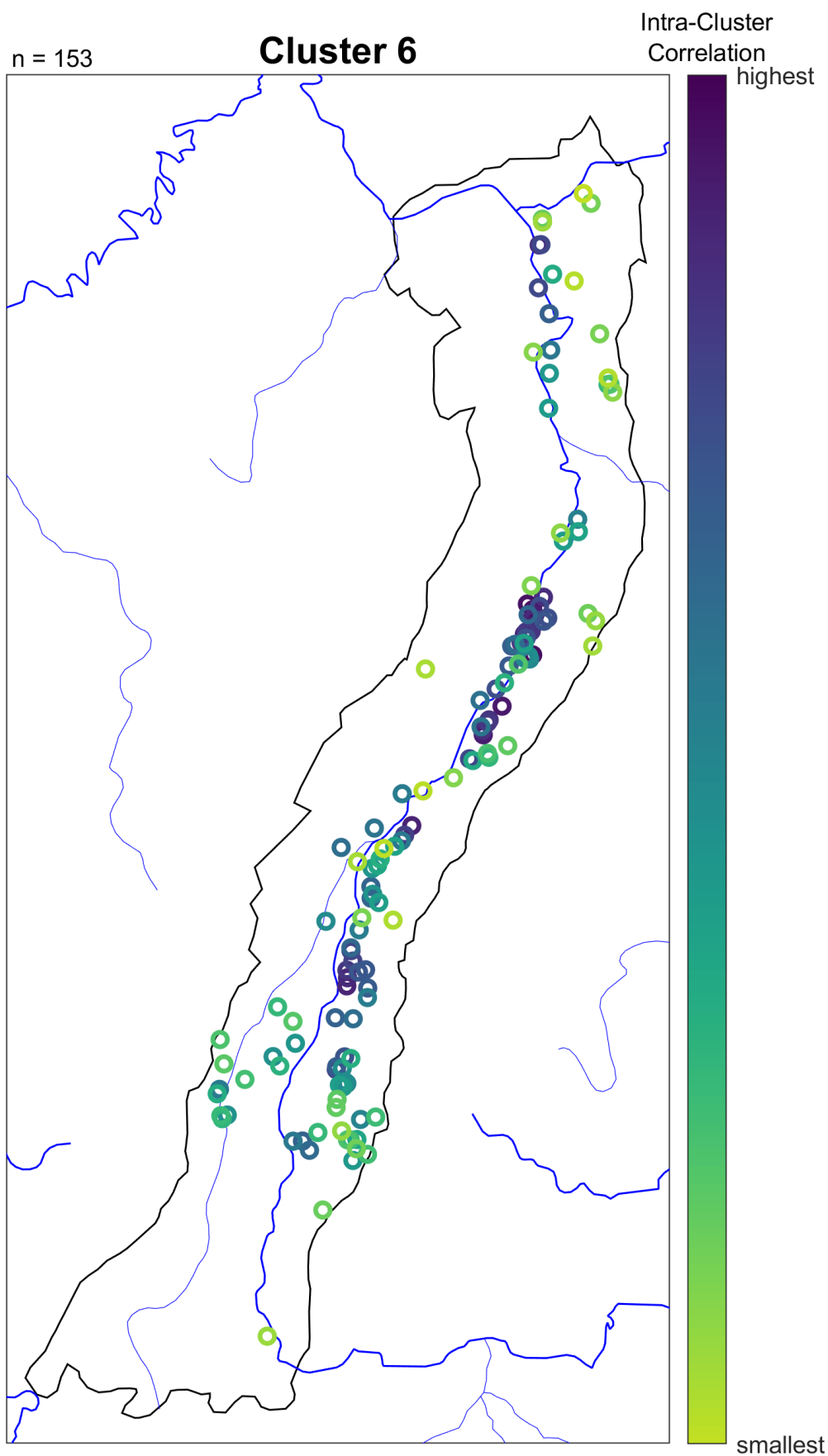
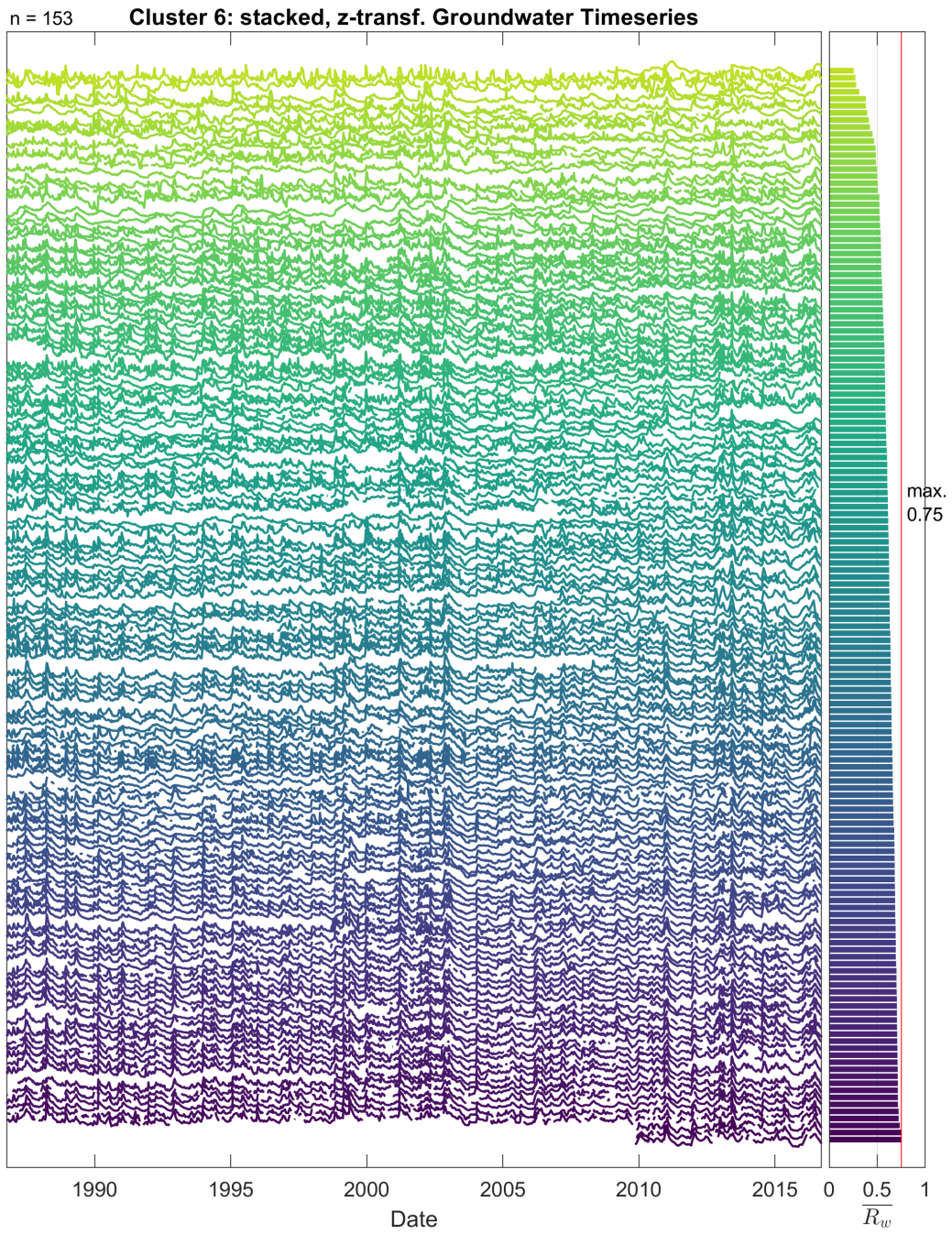


Fig. S41 Well locations in Cluster 6



**Fig. S42** Stacked, and z-scored hydrographs of Cluster 6

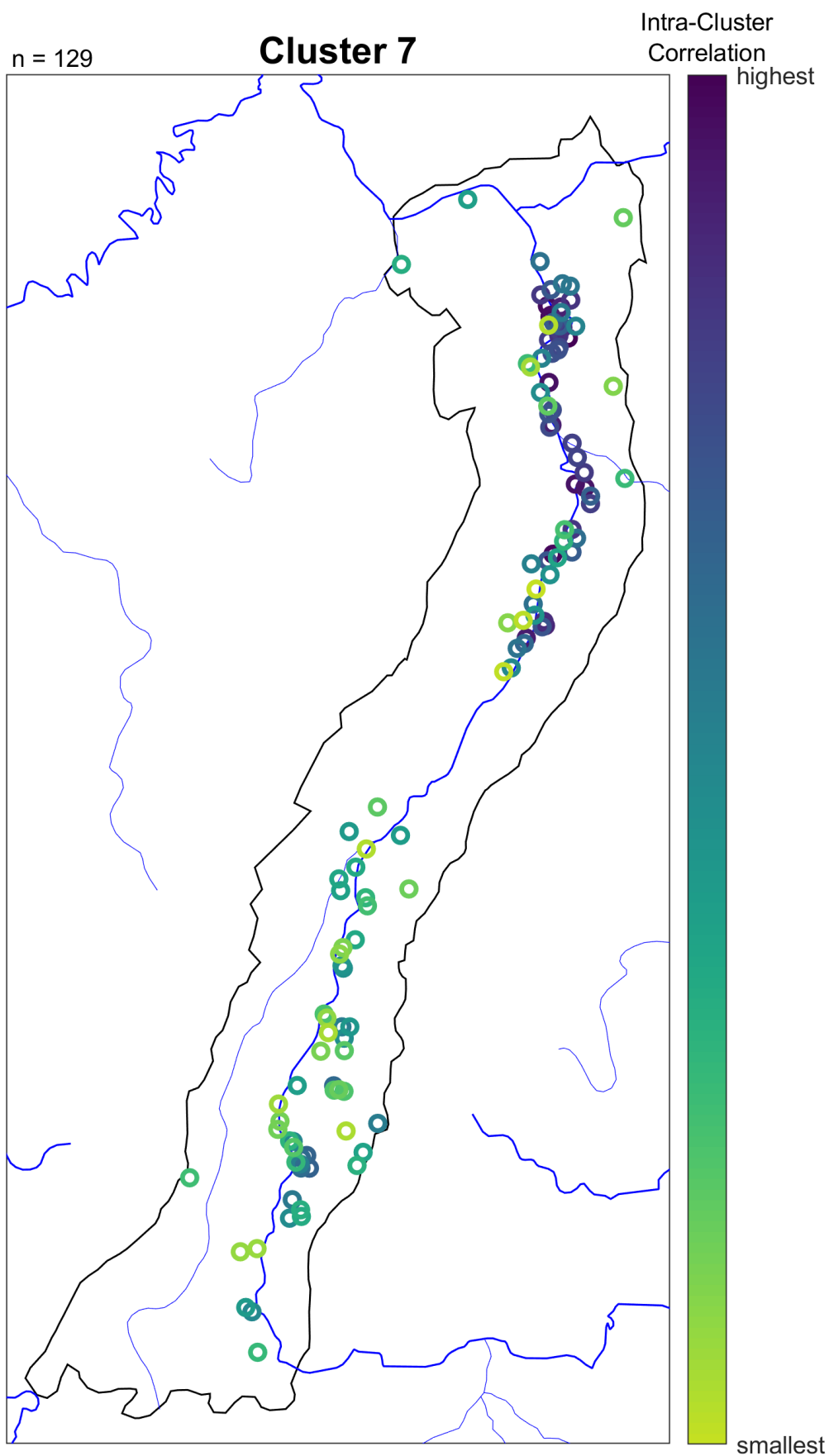
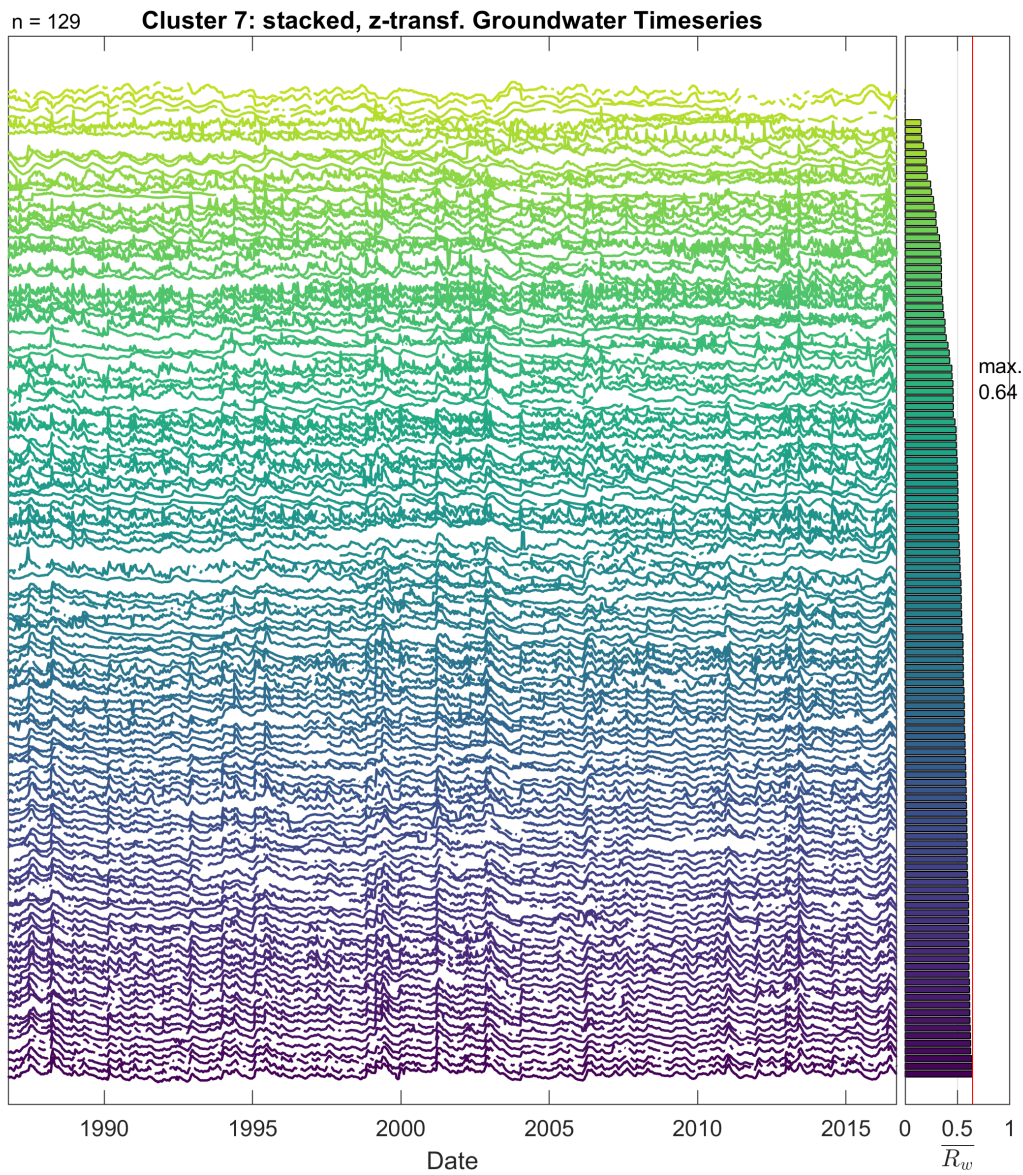


Fig. S43 Well locations in Cluster 7





**Fig. S44** Stacked, and z-scored hydrographs of Cluster 7

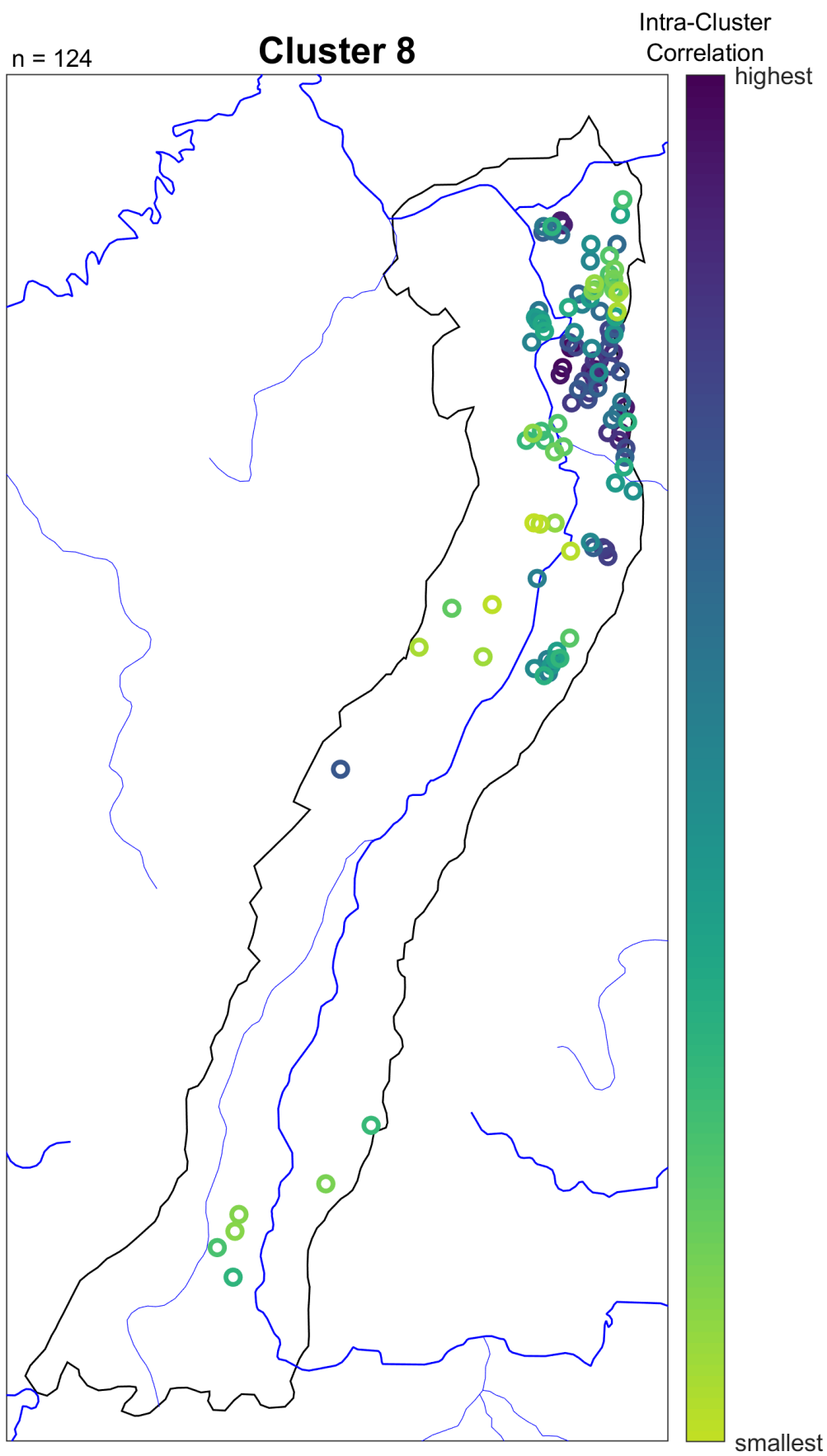
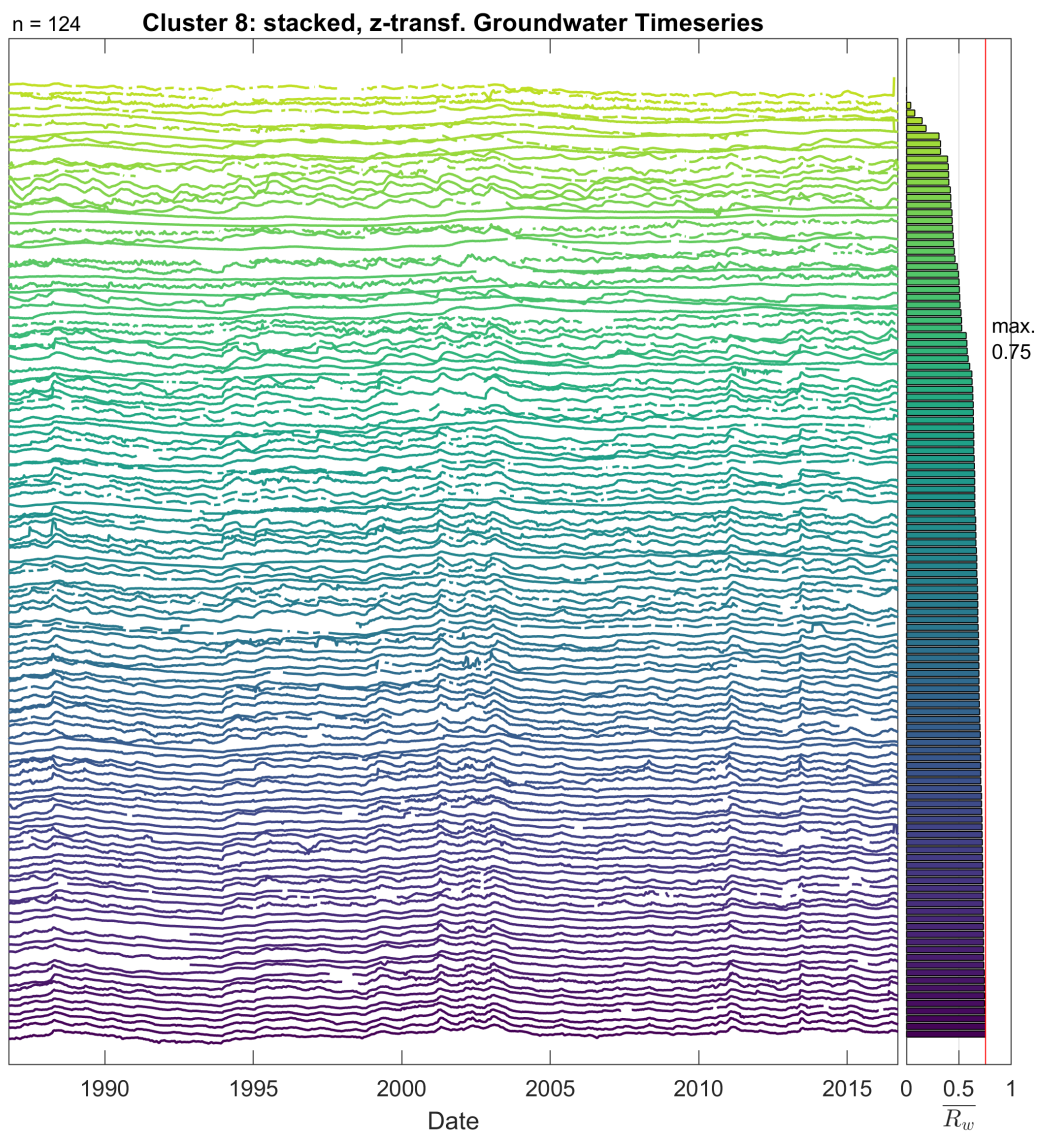


Fig. S45 Well locations in Cluster 8



**Fig. S46** Stacked, and z-scored hydrographs of Cluster 8

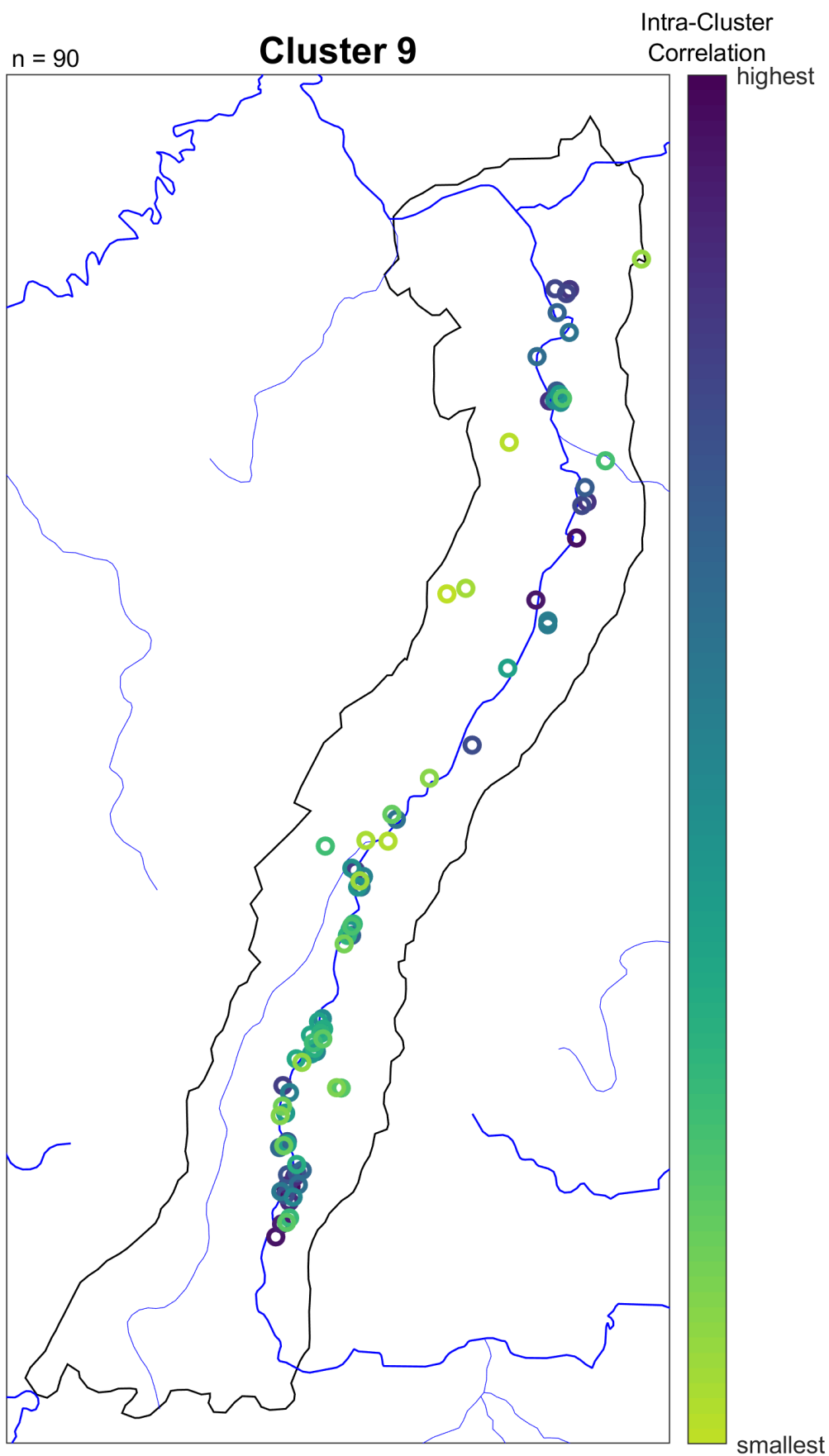
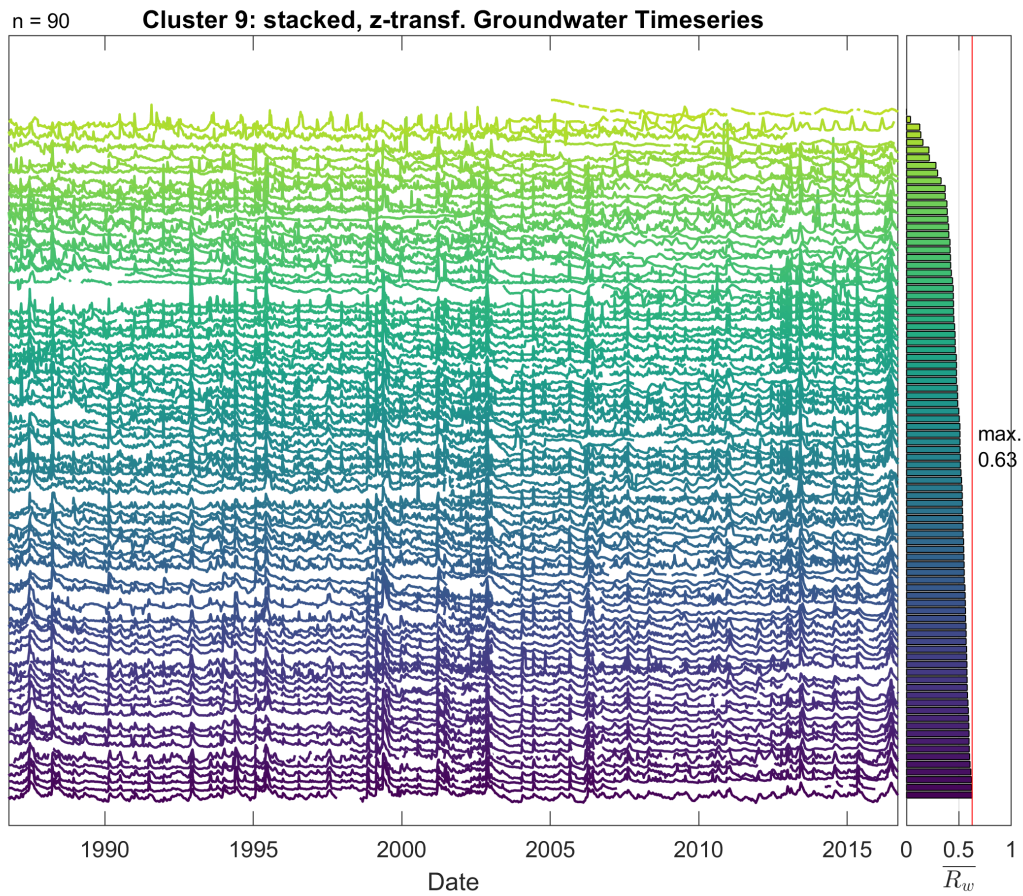


Fig. S47 Well locations in Cluster 9





**Fig. S48** Stacked, and z-scored hydrographs of Cluster 9



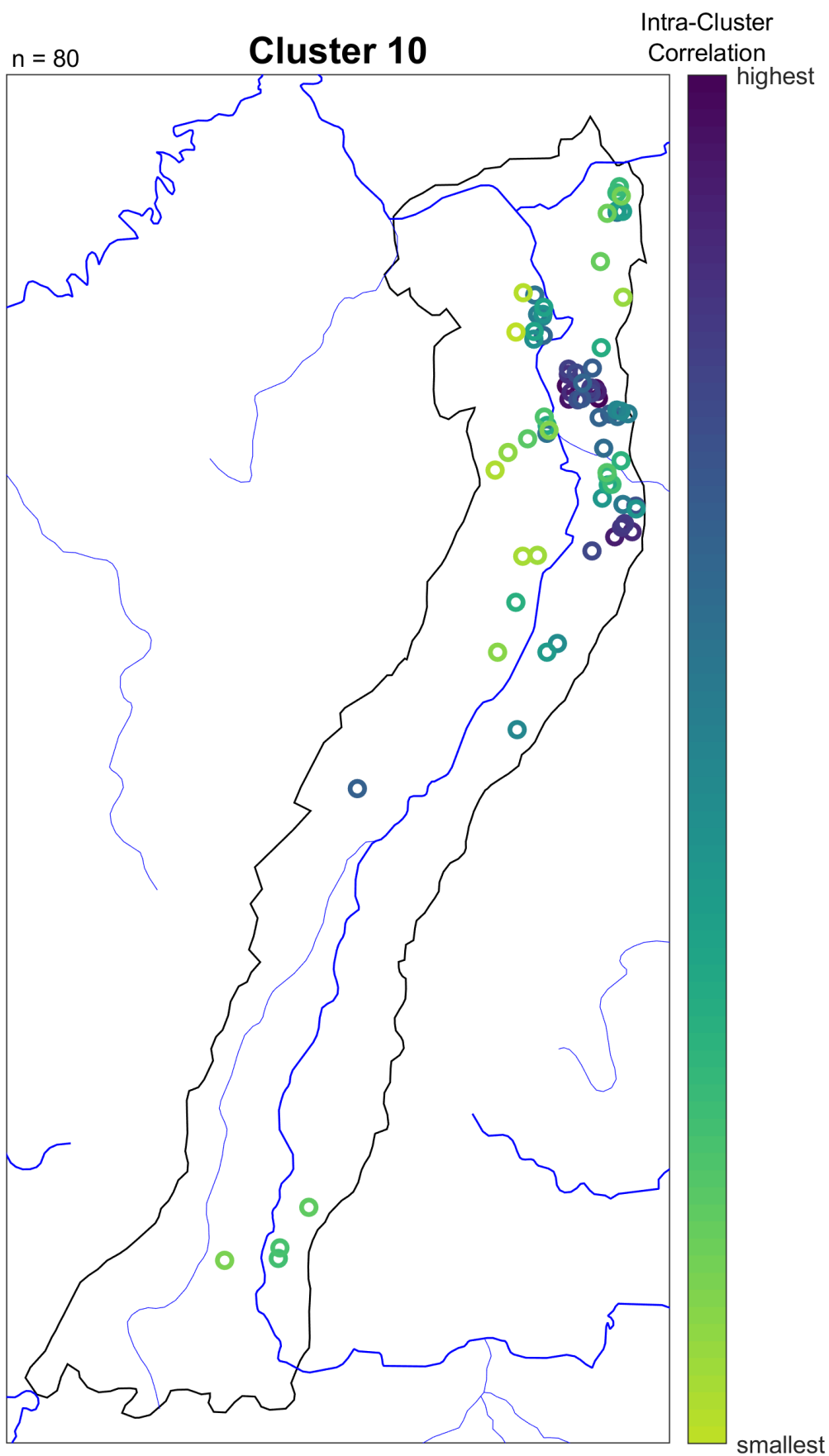
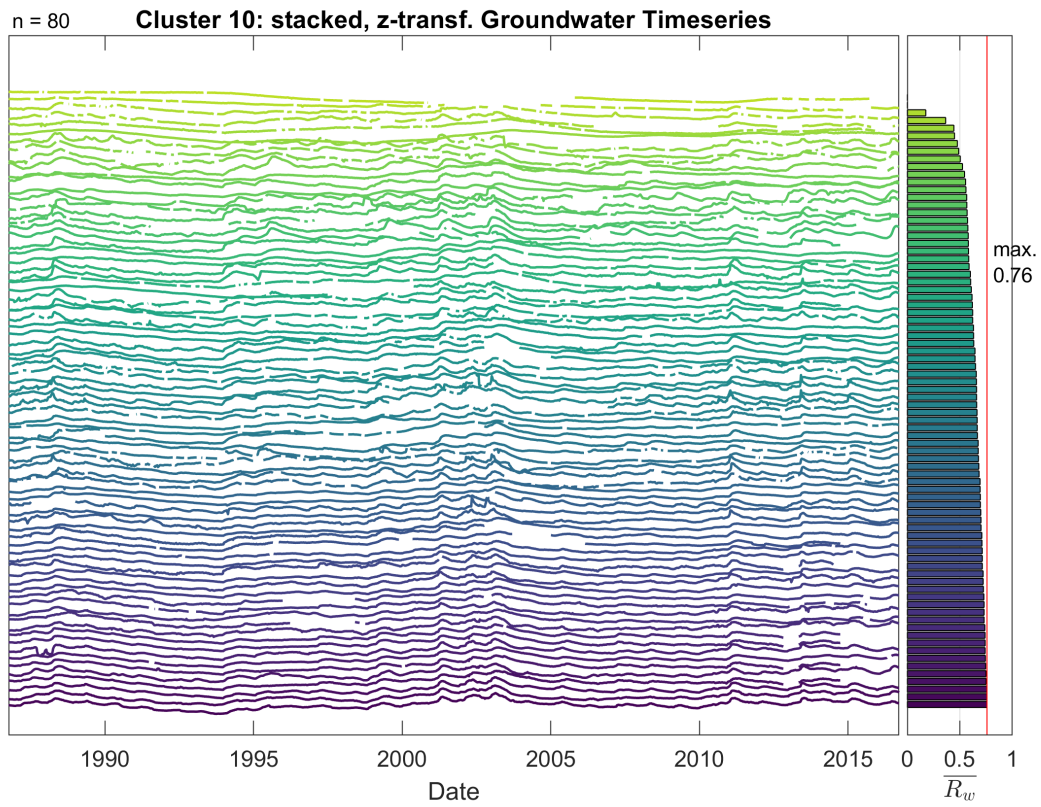


Fig. S49 Well locations in Cluster 10



**Fig. S50** Stacked, and z-scored hydrographs of Cluster 10

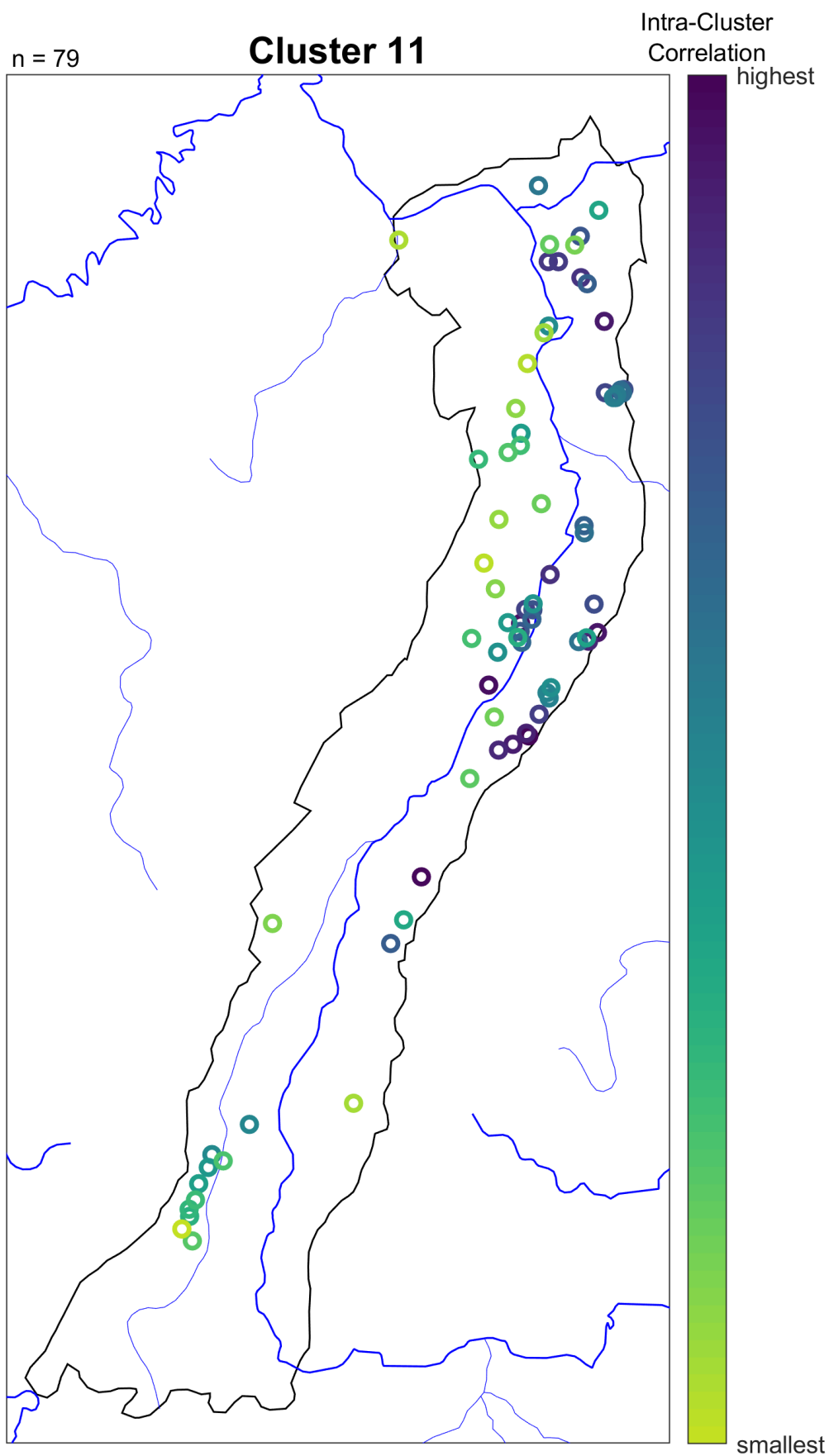
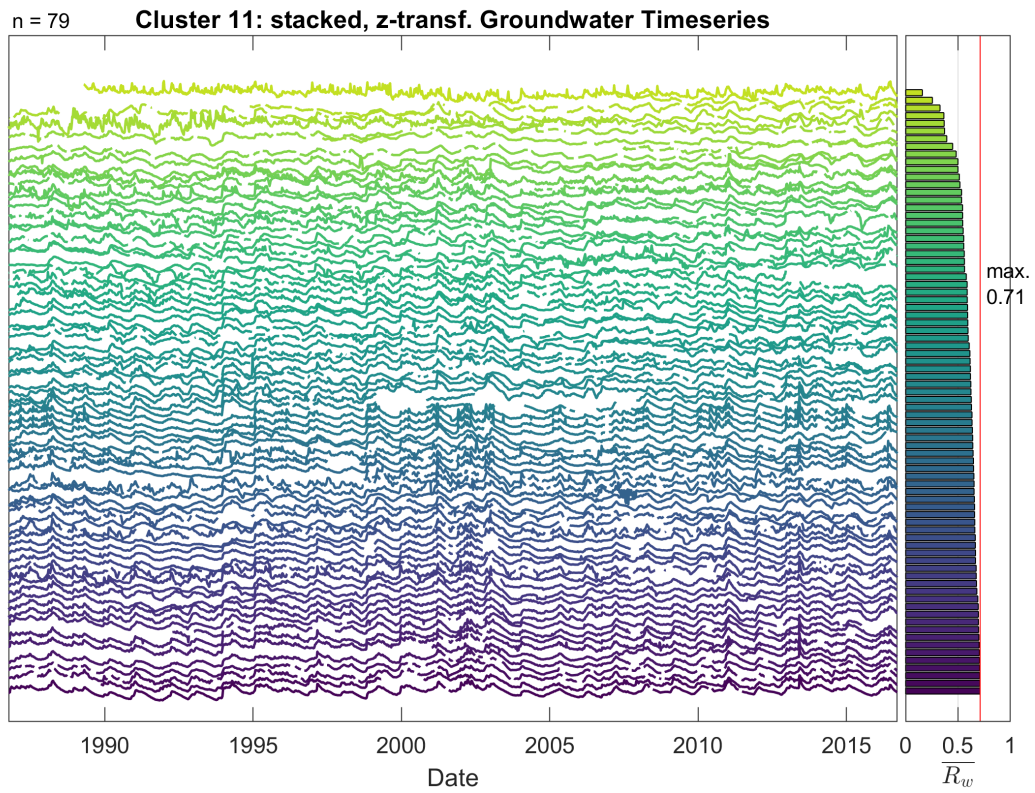


Fig. S51 Well locations in Cluster 11



**Fig. S52** Stacked, and z-scored hydrographs of Cluster 11

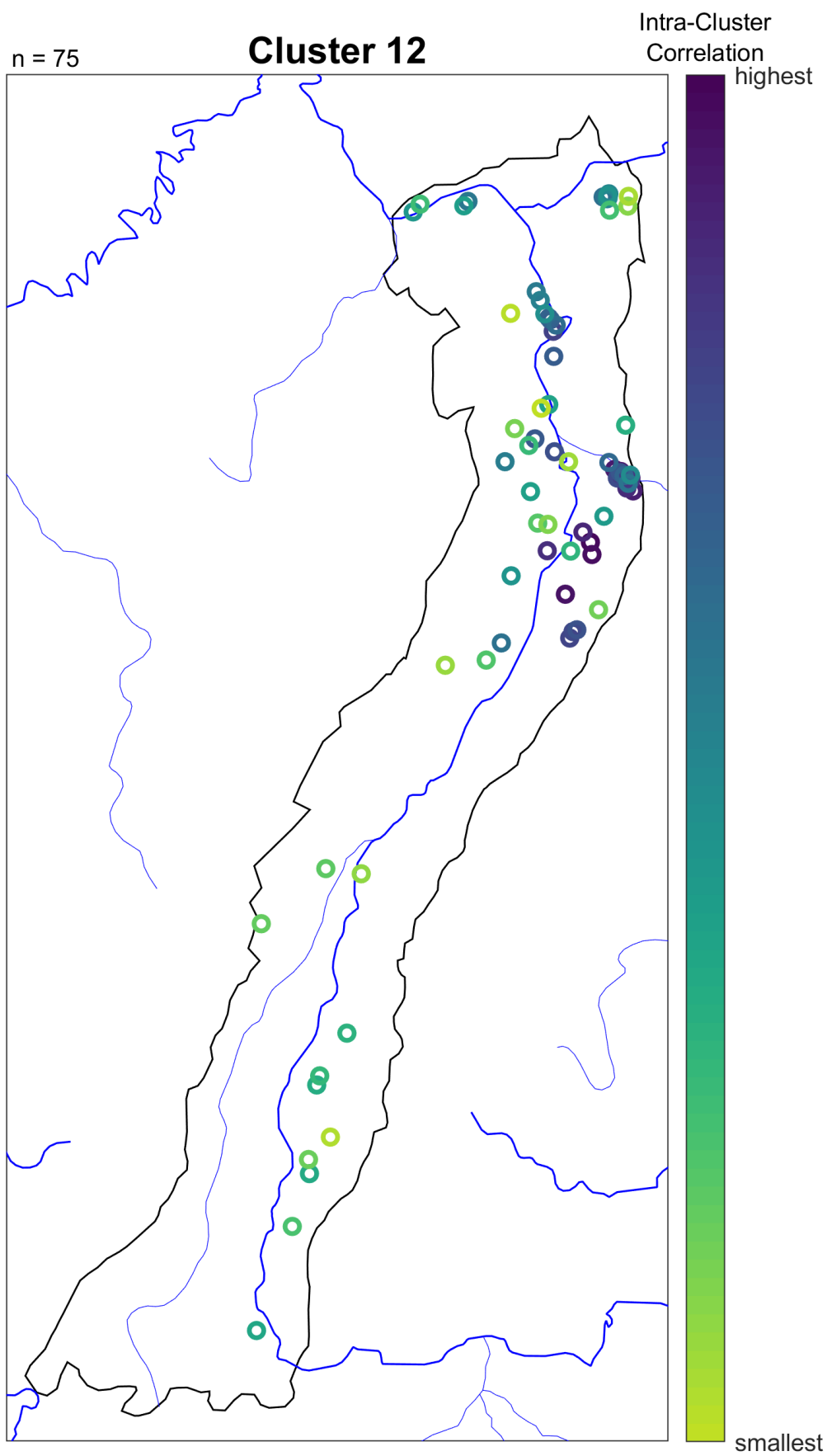
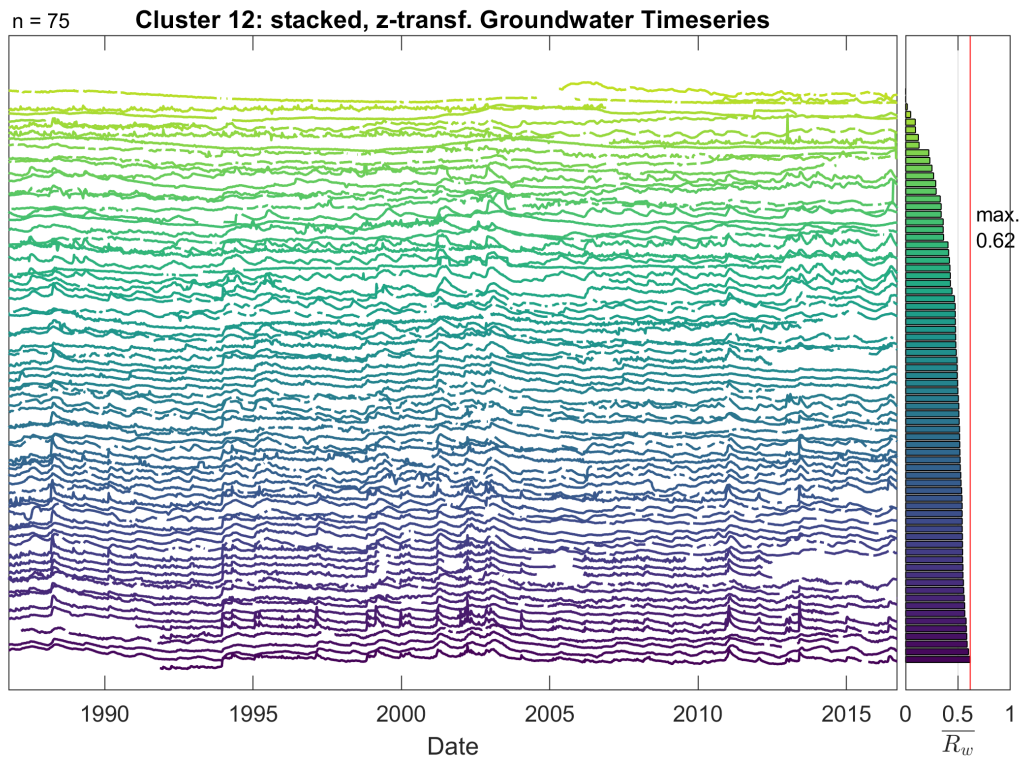


Fig. S53 Well locations in Cluster 12



**Fig. S54** Stacked, and z-scored hydrographs of Cluster 12

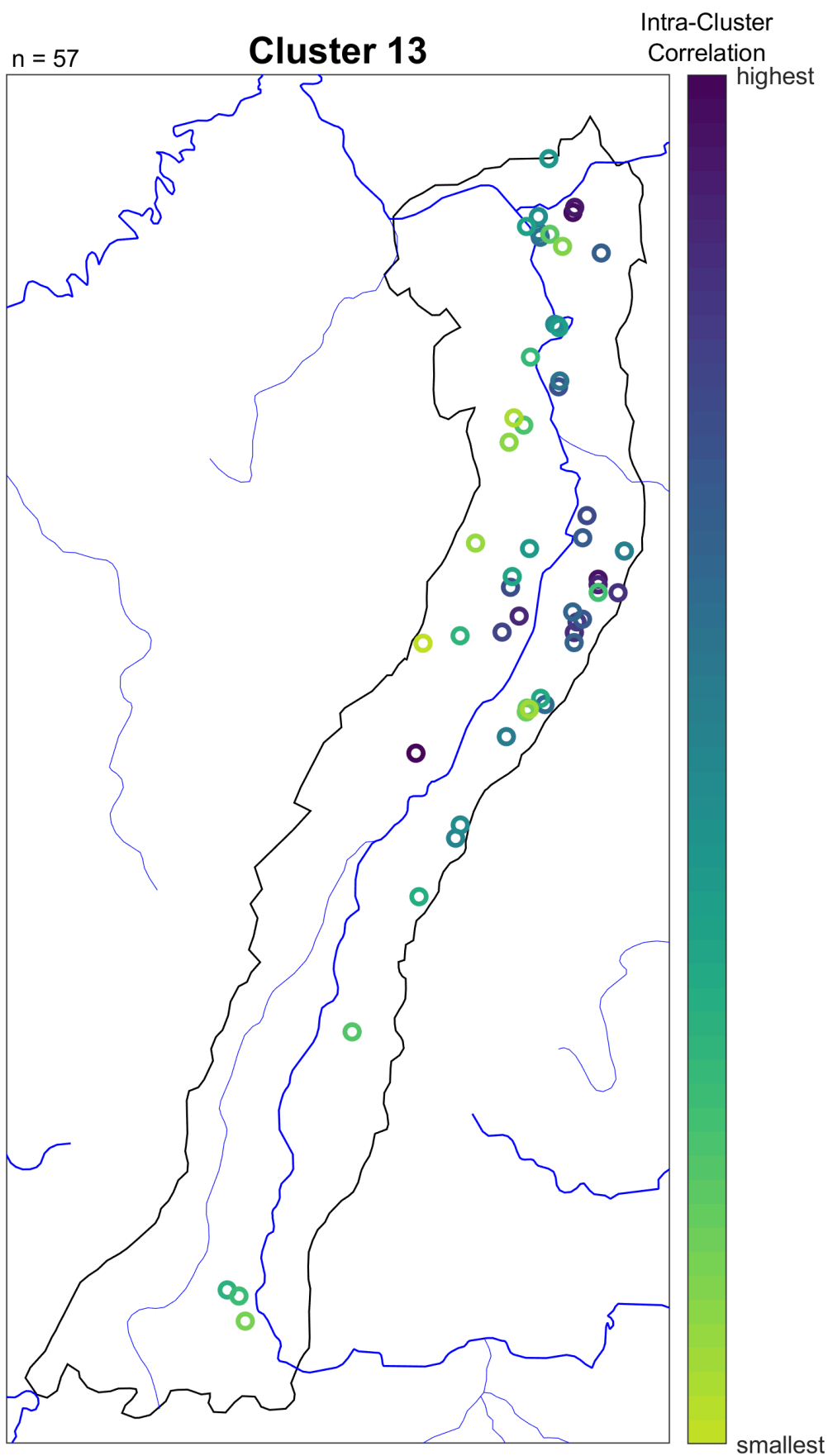
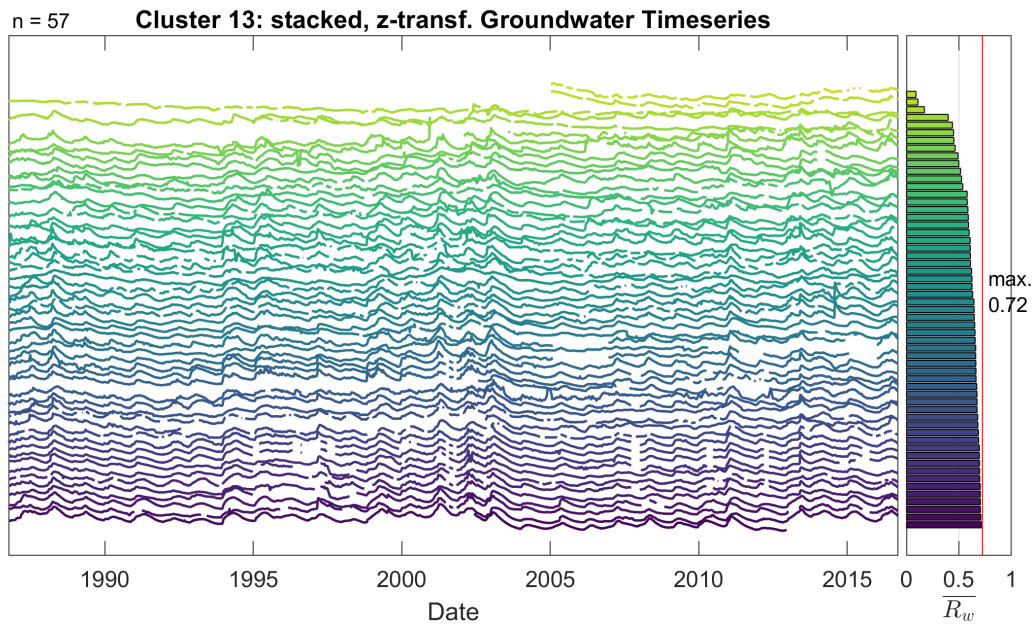


Fig. S55 Well locations in Cluster 13



**Fig. S56** Stacked, and z-scored hydrographs of Cluster 13



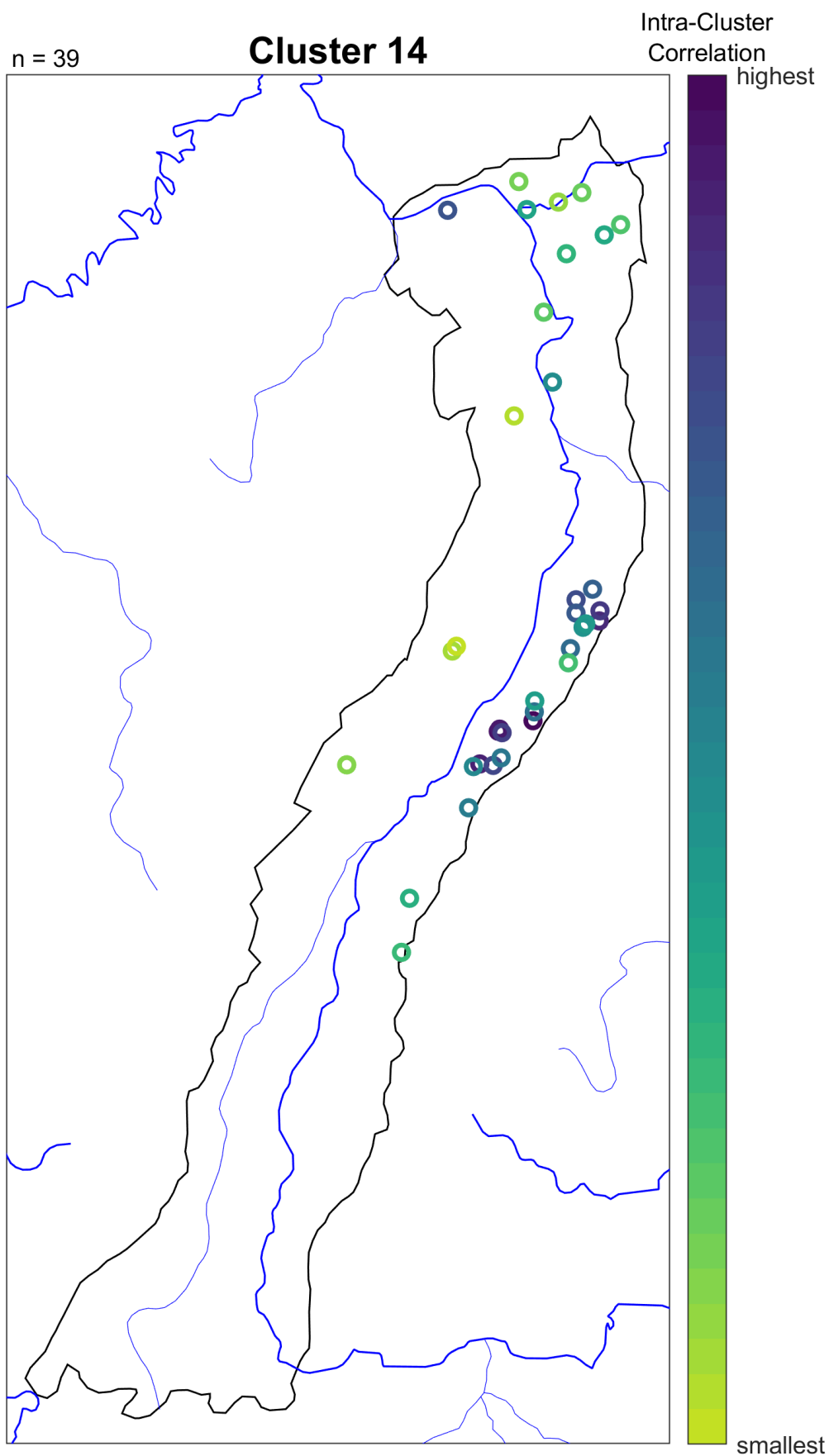
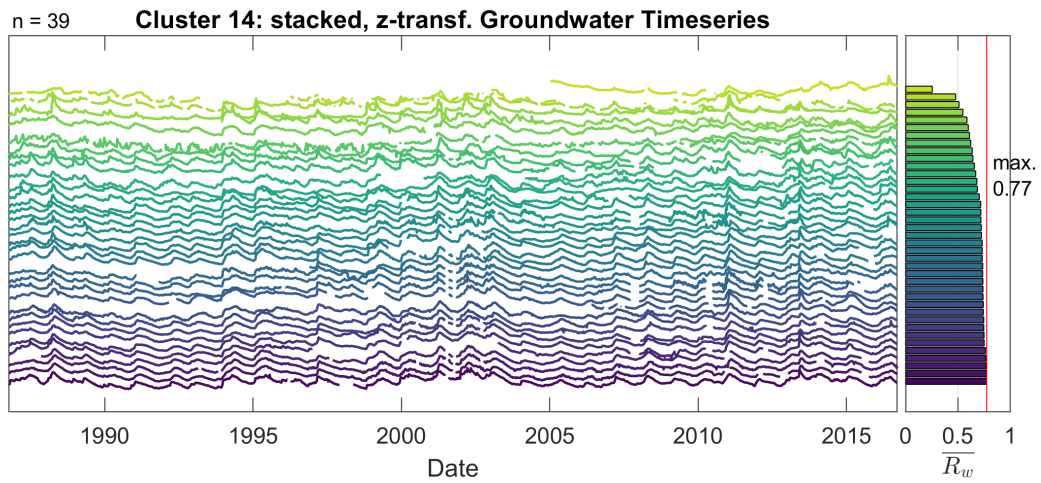
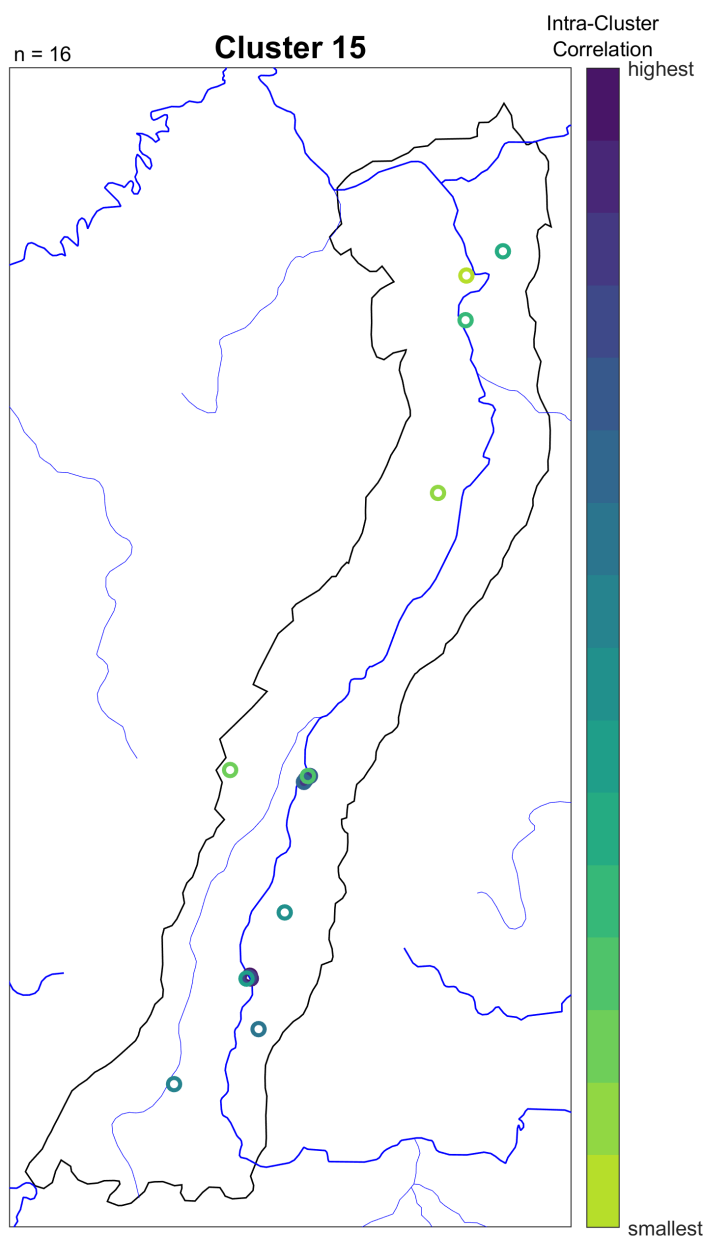


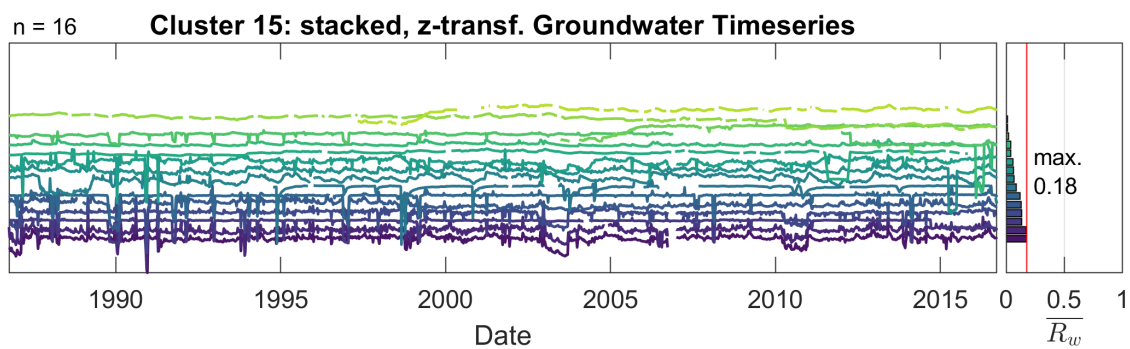
Fig. S57 Well locations in Cluster 14



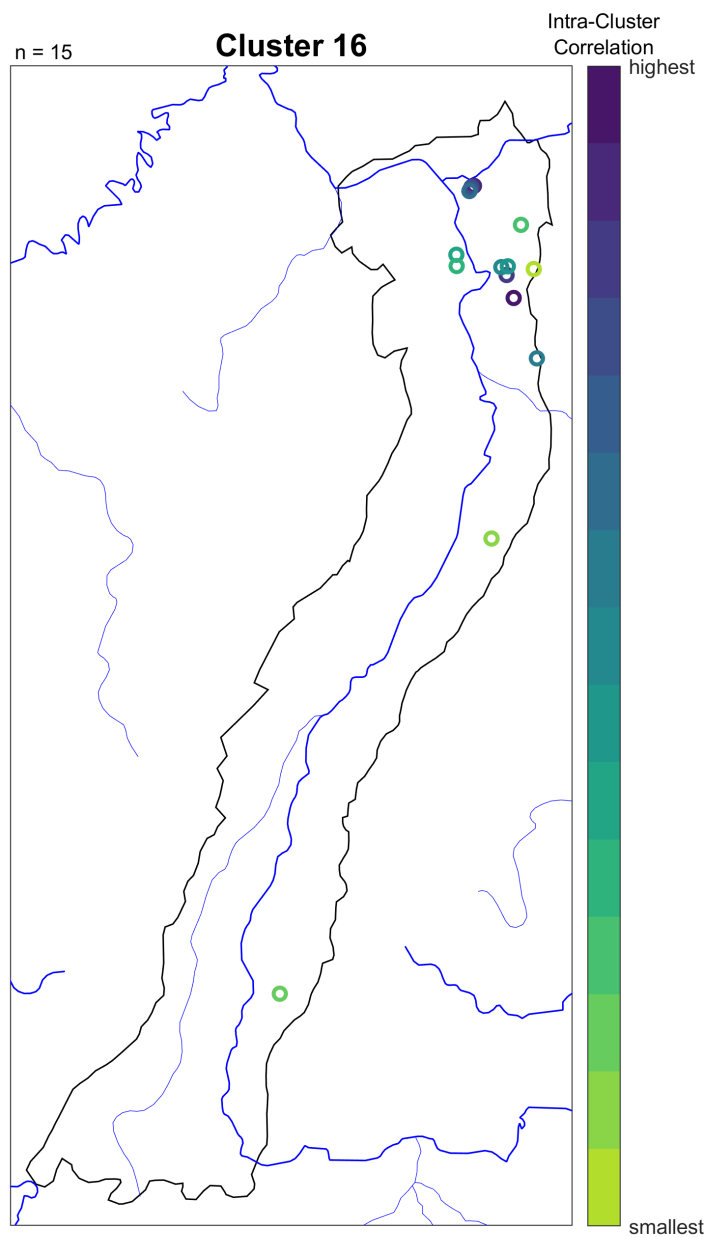
**Fig. S58** Stacked, and z-scored hydrographs of Cluster 14



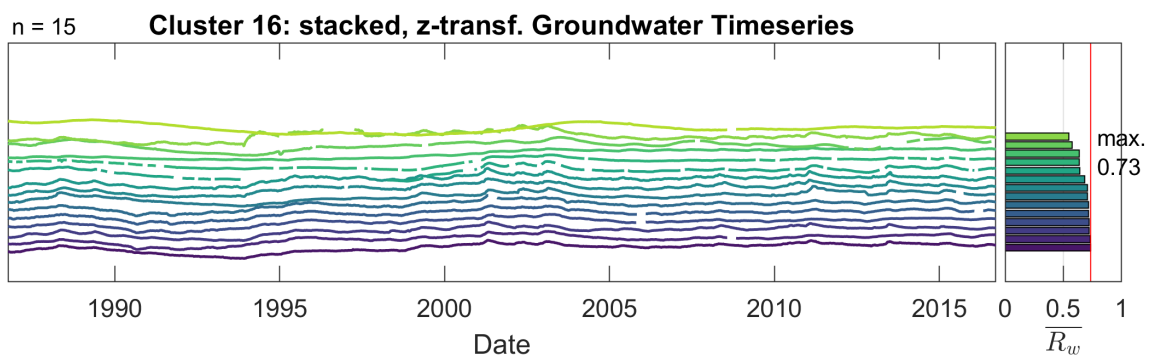
**Fig. S59** Well locations in Cluster 15



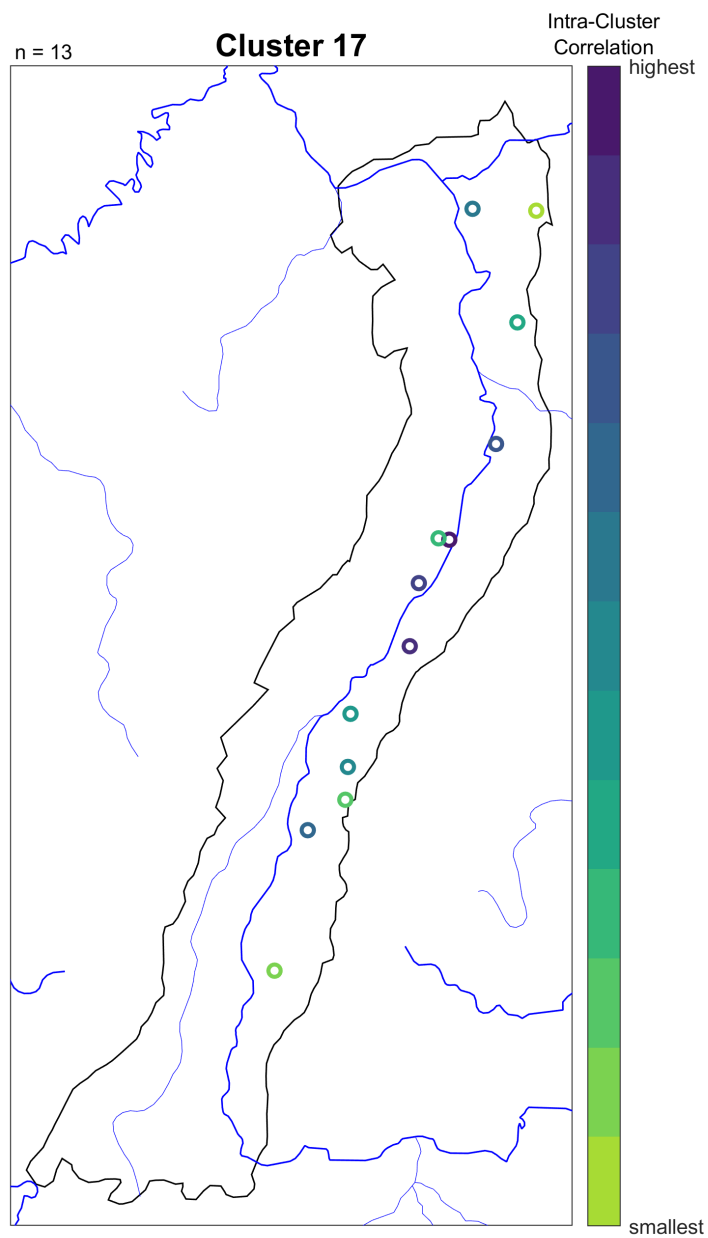
**Fig. S60** Stacked, and z-scored hydrographs of Cluster 15



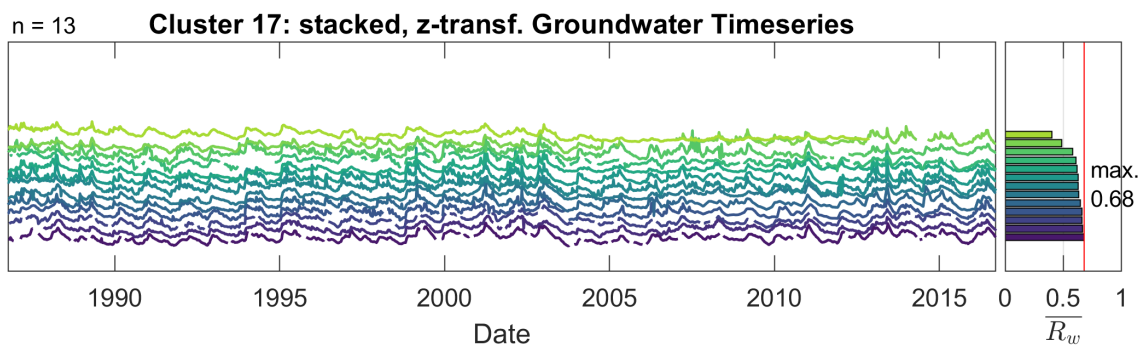
**Fig. S61** Well locations in Cluster 16



**Fig. S62** Stacked, and z-scored hydrographs of Cluster 16

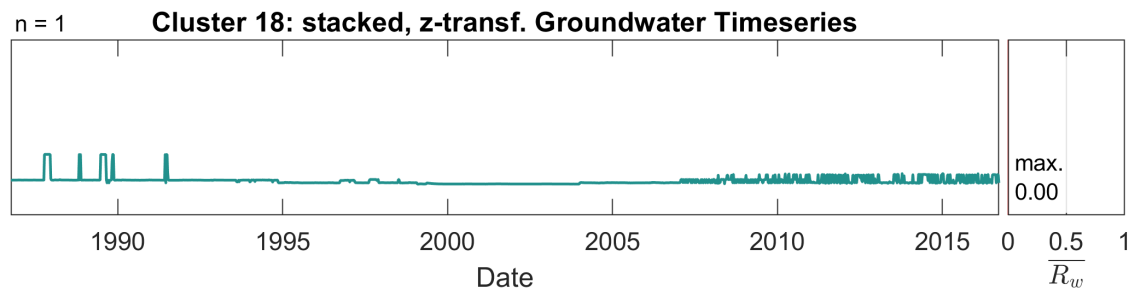


**Fig. S63** Well locations in Cluster 17



**Fig. S64** Stacked, and z-scored hydrographs of Cluster 17

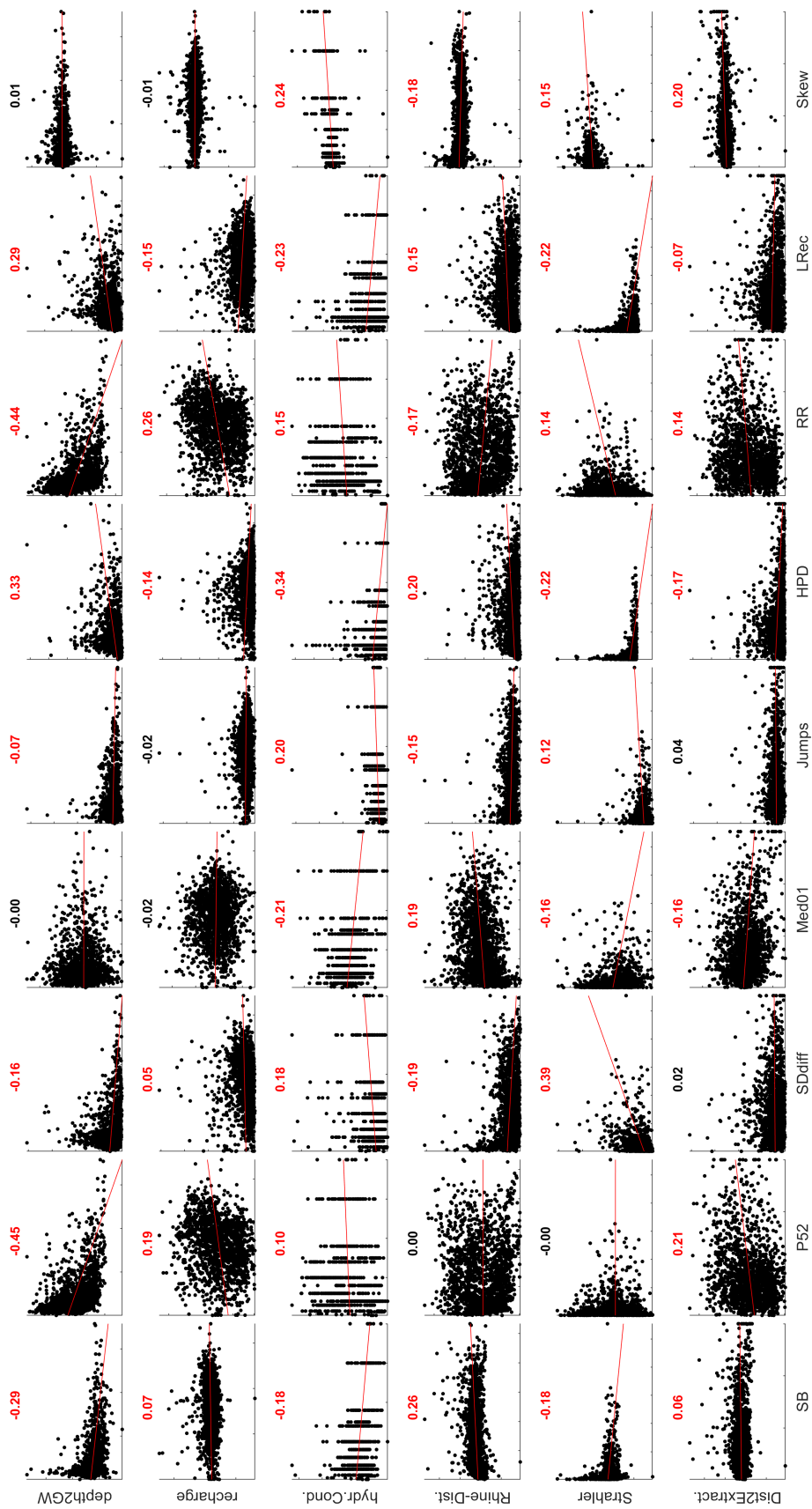




**Fig. S65** Z-scored hydrograph of Cluster 18

**Table S4** Summary of correlation analysis between features and influencing factors (n.s. - not significant)

Feature	RR	Skew	P52	SD <sub>diff</sub>	LRec	Jumps	SB	Med01	HPD	Comment
Mean Depth to GW	-0.44	n.s.	-0.45	-0.16	0.29	-0.07	-0.29	n.s.	0.33	
Diffuse GW-Recharge	0.26	n.s.	0.19	0.05	-0.15	n.s.	0.07	n.s.	-0.14	1663 german wells only
Hydr. Conductivity	0.15	0.24	0.1	0.18	-0.23	0.2	-0.18	-0.21	-0.34	Subset of german wells (828), Spearman Rank-Correlation
Streamflow-Influence (Dist. to Rhine River)	-0.17	-0.18	n.s.	-0.19	0.15	-0.15	0.26	0.19	0.2	Distance as surrogate for influence
Streamflow Influence (Strahler-Classes)	0.14	0.15	n.s.	0.39	-0.22	0.12	-0.18	-0.16	-0.22	Influence is considered directly
Distance to GW Extractions	0.06	0.2	0.21	n.s.	-0.07	n.s.	0.07	-0.16	-0.17	1663 german wells only



**Fig. S66** Summary of correlation analysis between features and influencing factors (black: not significant), hydraulic conductivity correlations are based on Spearman rank-correlation instead of linear Pearson correlation, due to categorical data.

## References

- Alley WM, Healy RW, LaBaugh JW, Reilly TE (2002) Flow and storage in groundwater systems. *science* 296(5575):1985–1990, DOI 10.1126/science.1067123
- Alqurashi TM, Wang W (2018) Clustering ensemble method. *International Journal of Machine Learning and Cybernetics* DOI 10.1007/s13042-017-0756-7
- Ayad HG, Kamel MS (2010) On voting-based consensus of cluster ensembles. *Pattern Recognition* 43(5):1943–1953, DOI 10.1016/j.patcog.2009.11.012
- Balugani E, Lubczynski M, Reyes-Acosta L, van der Tol C, Francés A, Metselaar K (2017) Groundwater and unsaturated zone evaporation and transpiration in a semi-arid open woodland. *Journal of Hydrology* 547:54–66, DOI 10.1016/j.jhydrol.2017.01.042
- BGR (2019) Mean Annual Groundwater Recharge of Germany 1:1,000,000 (GWN1000). <https://www.bgr.bund.de/had>
- Bouwer H (2002) Artificial recharge of groundwater: Hydrogeology and engineering. *Hydrogeology Journal* 10(1):121–142, DOI 10.1007/s10040-001-0182-4
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics* 3(1):1–27, DOI 10.1080/03610927408827101
- Cloutier CA, Buffin-Bélanger T, Larocque M (2014) Controls of groundwater floodwave propagation in a gravelly floodplain. *Journal of Hydrology* 511:423–431, DOI 10.1016/j.jhydrol.2014.02.014
- CORINE Land Cover (2018) CORINE Land Cover (CLC) — Copernicus Land Monitoring Service. <https://land.copernicus.eu/user-corner/publications/clc-flyer>
- Corona CR, Gurdak JJ, Dickinson JE, Ferré T, Maurer EP (2018) Climate variability and vadose zone controls on damping of transient recharge. *Journal of Hydrology* 561:1094–1104, DOI 10.1016/j.jhydrol.2017.08.028
- Cuthbert MO (2014) Straight thinking about groundwater recession. *Water Resources Research* 50(3):2407–2424, DOI 10.1002/2013wr014060
- Desgraupes B (2018) clusterCrit: Clustering Indices
- Ghosh J, Acharya A (2011) Cluster ensembles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1(4):305–315, DOI 10.1002/widm.32
- Höltling B, Coldewey WG (2013) *Hydrogeologie: Einführung in die allgemeine und angewandte Hydrogeologie*, 8th edn. Springer-Spektrum, Berlin
- Hubert L, Schultz J (1976) Quadratic Assignment as a General Data Analysis Strategy. *British Journal of Mathematical and Statistical Psychology* 29(2):190–241, DOI 10.1111/j.2044-8317.1976.tb00714.x
- Jasechko S, Birks SJ, Gleeson T, Wada Y, Fawcett PJ, Sharp ZD, McDonnell JJ, Welker JM (2014) The pronounced seasonality of global groundwater recharge. *Water Resour Res* 50(11):8845–8867, DOI 10.1002/2014WR015809
- Lam A, Karssenberg D, van den Hurk BJJM, Bierkens MFP (2011) Spatial and temporal connections in groundwater contribution to evaporation. *Hydrol Earth Syst Sci* 15(8):2621–2630, DOI 10.5194/hess-15-2621-2011
- 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 Resources Research* 43(4), DOI 10.1029/2006wr005000
- 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
- McClain JO, Rao VR (1975) CLUSTISZ: A Program to Test for the Quality of Clustering of a Set of Objects. *Journal of Marketing Research* 12(4):456–460
- Pakhira MK, Bandyopadhyay S, Maulik U (2004) Validity index for crisp and fuzzy clusters. *Pattern Recognition* 37(3):487–501, DOI 10.1016/j.patcog.2003.06.005
- Ratkowsky D, Lance G (1978) A criterion for determining the number of groups in a classification. *Australian Computer Journal* 10(3):115–117
- Regierungspräsidium Darmstadt (1999) *Grundwasserbewirtschaftungsplan Hessisches Ried*. Tech. rep.
- Shestakov A (2017) Consensus clustering experiment framework
- Sinharay S (2010) Jackknife Methods. In: Peterson P, Baker E, McGaw B (eds) *International Encyclopedia of Education (Third Edition)*, Elsevier, Oxford, pp 229–231, DOI 10.1016/B978-0-08-044894-7.01338-5
- Stoll S, Hendricks Franssen HJ, Barthel R, Kinzelbach W (2011) What can we learn from long-term groundwater data to improve climate change impact studies? *Hydrol Earth Syst Sci* 15(12):3861–3875,

- DOI 10.5194/hess-15-3861-2011
- Thierion C, Longuevergne L, Habets F, Ledoux E, Ackerer P, Majdalani S, Leblois E, Lecluse S, Martin E, Queguiner S, Viennot P (2012) Assessing the water balance of the Upper Rhine Graben hydrosystem. *Journal of Hydrology* 424-425:68–83, DOI 10.1016/j.jhydrol.2011.12.028
- Toth E (2009) Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. *Hydrol Earth Syst Sci* p 12, DOI 10.5194/hess-13-1555-2009
- Vega-Pons S, Ruiz-Shulcloper J (2011) A survey of clustering ensemble algorithms. *Int J Patt Recogn Artif Intell* 25(03):337–372, DOI 10.1142/s0218001411008683
- Vesanto J (2000) SOM Toolbox for Matlab 5. Helsinki University of Technology, Espoo