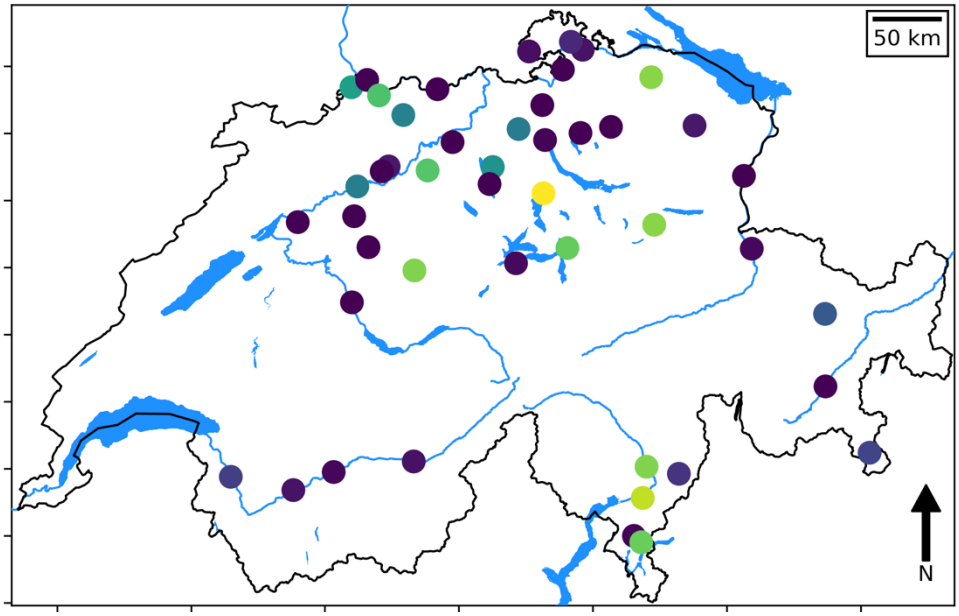
Groundwater signatures and spatial clustering of the NAQUA-QUANT stations



Dübendorf, 2024.12.10

Commissioned by the Federal Office for the Environment (FOEN)

Imprint

Commissioned by: Federal Office for the Environment (FOEN), Hydrology Division, CH 3003 Bern. The FOEN is an agency of the Federal Department of the Environment, Transport, Energy and Communications (DETEC).
Contractor: Eawag
Author: Dr. R.A. Collenteur (Eawag)
FOEN support: Dr. J. von Freyberg
Note: This study/report was prepared under contract to the Federal Office for the Environment (FOEN). The contractor bears sole responsibility for the content.

Table of Contents

1 Introduction

In this report, the hydraulic head time series from the NAQUA QUANT monitoring network is analyzed. The study consists of two parts. The goal of the first part is to test the standardized groundwater index (SGI) as an indicator of groundwater drought events, compared to historic events. The goal of the second part is to use groundwater signatures to cluster different groundwater monitoring stations into smaller groups with similar behavior.

This report is a complement to a delivery package with data and Jupyter Notebooks (readable Python scripts) of the analysis, which are the main output of this project. At the end of each chapter, a brief description of the used Jupyter Notebook, the input and output data, and the produced figures is provided. This report concisely summarizes the main results of this study.

This report is structured as follows. In the second chapter, the data used for the analysis is summarized and described. The third chapter briefly describes the modeling of the hydraulic head time series and how the models are used to gap-fill the head time series. The fourth chapter describes the analysis of the SGI and how it compares to the current FOEN approach based on percentiles. The fifth chapter describes the results of the clustering analysis using groundwater signatures. The sixth and final chapter summarizes this report.

2 Data description

2.1 Hydraulic head time series

Time series of hydraulic heads (hereafter, heads) from 51 pumping wells and piezometers were provided by the FOEN from the QUANT module. For the initial screening, the data was plotted in graphs and visually inspected for any clear errors in the data that could impact the remainder of the analysis. After this initial inspection, the data from five monitoring stations were updated by the FOEN and new data were delivered and used in this study. As a result, all of the 51 stations could be used for further analysis.

For many statistical analyses, it is important to have continuous, gapless time series over the period of interest. Here, the 30-year period 1993-2023 was chosen as the period of interest. This is in line with the 30-year period commonly used in climate studies for statistical (trend) analysis. Moreover, as the meteorological data is available from 1990 onwards, this allows for a three-year warm-up period for the hydrological model used to simulate the hydraulic heads.

From the 51 stations, 39 head time series had missing data. The minimum and maximum percentage of missing data varied between 0 and 65% of the daily head values, with an average percentage of 13%. These results are visualized in Figures 1 and 2. Figure 1 shows the locations of the monitoring stations and the percentage of missing data. From the data shown in Figure 1, it can be observed that there is no clear spatial pattern in where stations with missing data are located.

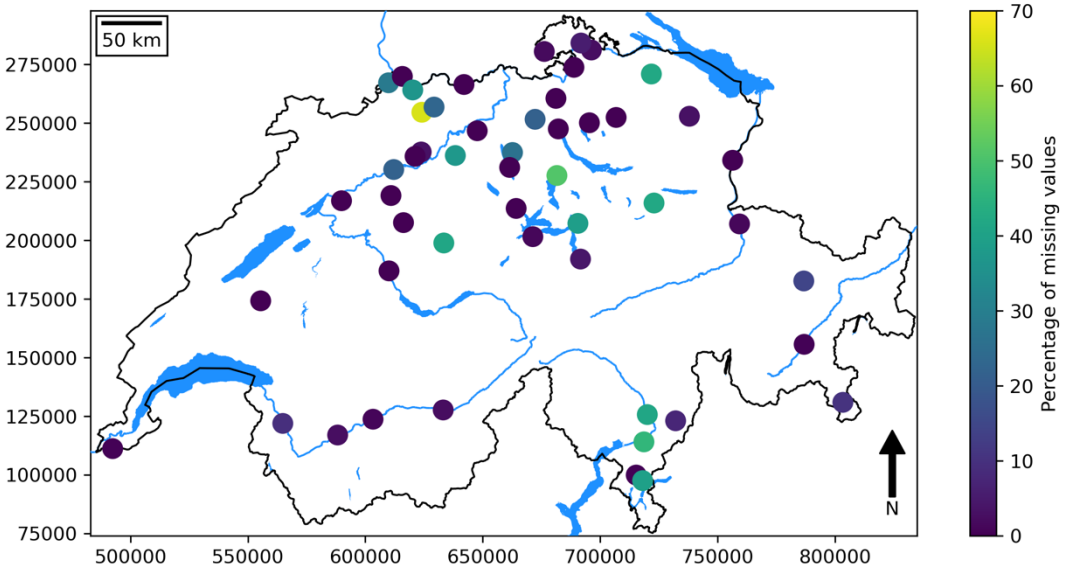


Figure 1. *Map of the locations of the monitoring stations used for the analysis. The color denotes the percentage of missing data in the daily time series from 1993-2023.*

Figure 2 shows the temporal data availability of the time series. The color denotes the number of observations per month, and it can be seen that all stations provide head data with a daily measurement frequency. Note that higher frequency data might be available, but that this was not used in this study. The data in Figure 2 also indicates that the largest part of the data

gaps is at the beginning of the period of interest, simply because the measurement started later (at 20 stations). For all the time series with missing data, gap-filling is applied to obtain 30-year gapless time series.

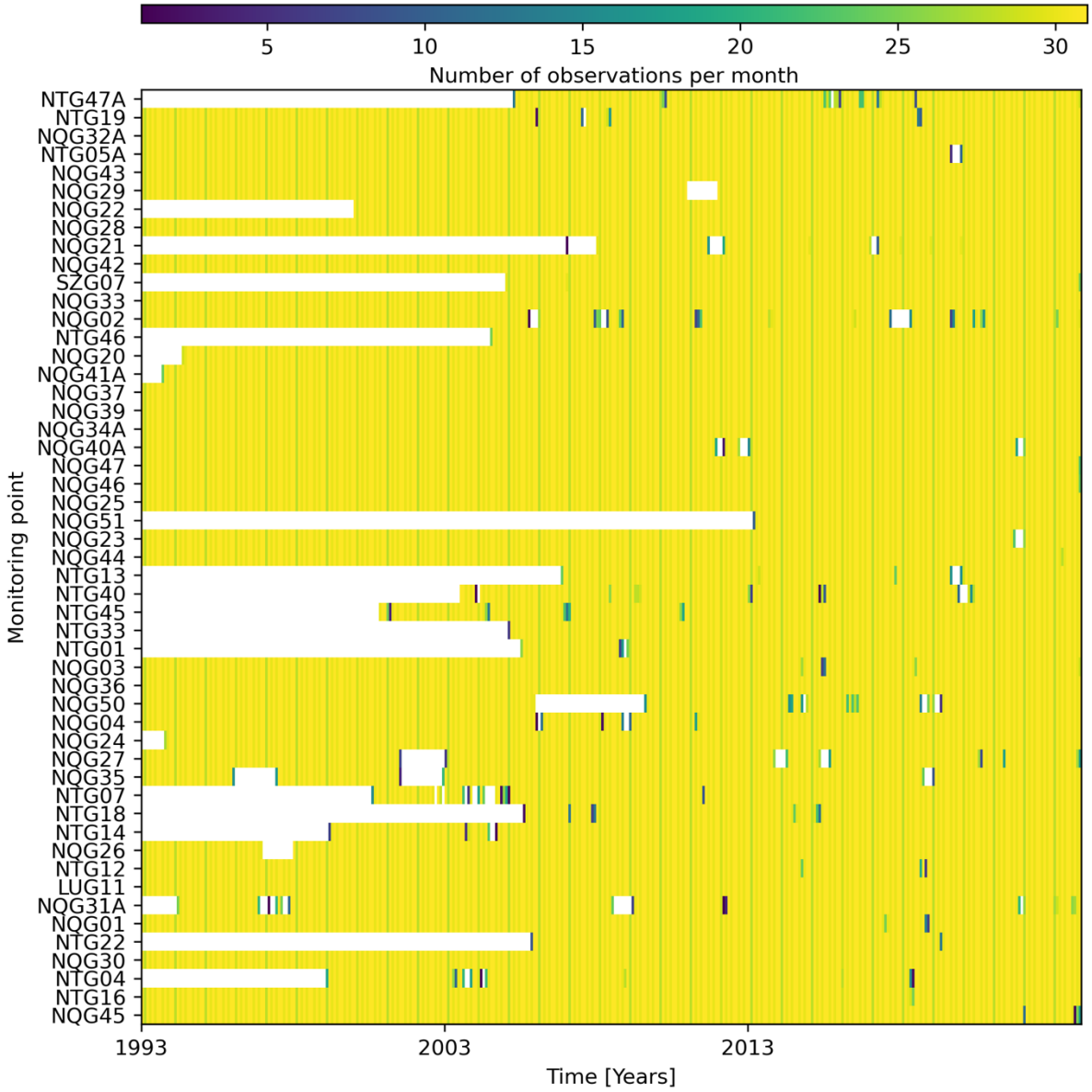


Figure 2. Visualization *of the data availability of the groundwater level data. Note that all the time series have daily values, most with only a few gaps. At 20 stations, measurements started (well) after 1993, causing a 'gap' at the start of the period of interest (1993-2003).*

2.2 Spring discharge data

In total, 43 time series of spring discharge were provided by the FOEN from the QUANT module. Out of these, 41 time series had data gaps. The percentage of missing data ranges between 0% and 100% of the data (station Cormoret), with an average of 42%. This average percentage of missing data is substantially higher compared to the head data (13%). Given this high percentage of missing spring discharge data, gap-filling is much more challenging. Rudolph et al. (2023) tested an adapted version of Pastas on spring discharge data with

moderate success, but fundamentally the Pastas model was not designed for this purpose and data. It was therefore decided not to gap-fill and/or use the spring discharge data for further analysis within this project. It is recommended to explore the use of other (types of) models to gap-fill the spring discharge data, for example neural network models.

2.3 Meteorological data

Meteorological data for the locations of the monitoring stations for the hydraulic heads is obtained from MeteoSwiss. Daily precipitation sums are taken from the gridded RhiresD dataset, and daily average air temperature from the gridded TabsD dataset for the period January 1991 to September 2023 (MeteoSwiss, 2023). Individual precipitation and air temperature time series for each groundwater monitoring station are extracted by selecting the data from the cell in which the monitoring station is located. Potential evaporation (PET) was computed from the air temperature data using the Hamon method implemented in the PyET Python package (Vremec et al., 2024).

2.4 Script and output data description

The script to preprocess the hydraulic head data is called *00_read_groundwater_data.ipynb*. The script uses the data from the '*raw*' folder, where all the data delivered by the FOEN is located. The script reads all raw data, and stores it in a format that can be more easily used for data analysis and modelling. This preprocessed data is saved into the '*processed*' folder. The head time series are stored in '*heads.csv*' and the metadata in the '*metadata_heads.csv*'. This script collects all this data and makes plots of the time series that are used for a visual inspection.

The meteorological data is collected and processed in the script called *01_get_meteo_data.ipynb*, using RhiresD and TabsD data from MeteoSwiss available at the Eawag. This script cannot be run without this data, which have to be requested from MeteoSwiss. The processed precipitation, air temperature and PET data per monitoring station, however, is saved and available from the '*processed*' folder. The precipitation, temperature, and potential evaporation are stored in '*precipitation.csv'*, '*temperature.csv'*, and '*evaporation.csv'*, respectively.

3 Gap-filling the head time series

3.1 Hydraulic head modelling

The original head time series were gap-filled using heads simulated with a lumped-parameter groundwater model from the Pastas modelling software (Collenteur et al. 2019). For each monitoring station, a separate model was constructed and calibrated. The heads were simulated using precipitation, potential evaporation, and temperature as model inputs. Other stresses on the groundwater system causing water table fluctuations were not considered, limiting the performance of the model for some locations. This is particularly the case for those stations impacted by pumping and river level fluctuations. It is stressed here, however, that this is a data availability issue rather than a model deficiency, i.e., the model can take these stresses into account, but the input data are more difficult to obtain. Including other stresses than meteorology was outside the scope of this study.

The models were calibrated on all available head data for each monitoring station, by minimizing the sum of the squared residuals. The residuals are computed as the simulated minus the observed heads. The model performance is evaluated using the coefficient of determination (R^2), where a score of 1 indicates as a perfect model (i.e., simulated and observed heads are equal), and a score below 0 indicates that the mean head would be a better estimator.

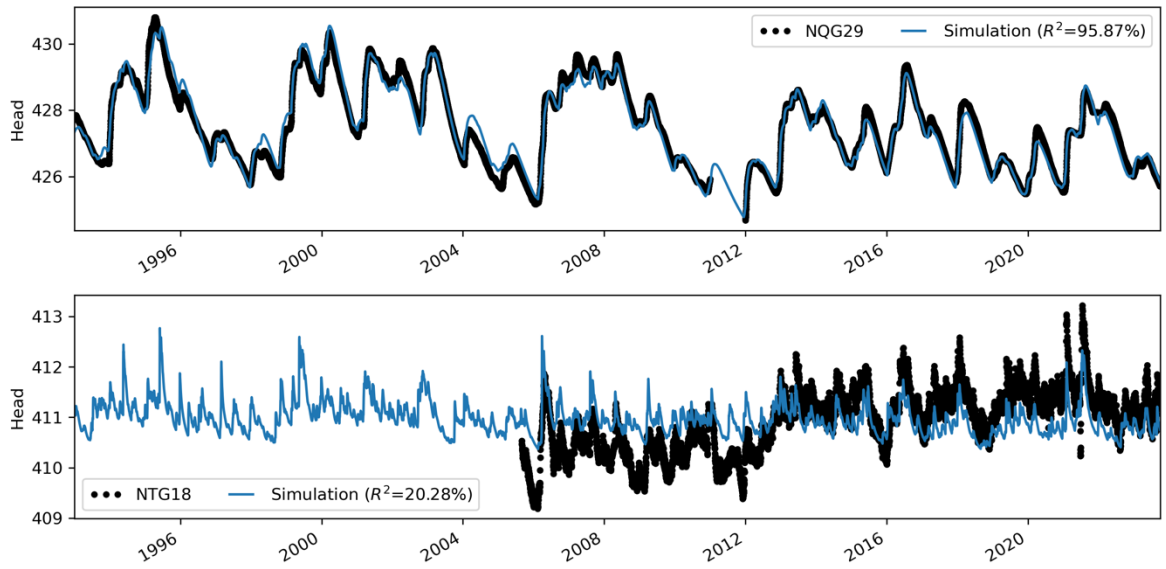


Figure 3. *Two examples of simulated and observed time series. The top panel shows a model with good performance (6535 Kestenholz), while the bottom panel shows a model with lower performance (6569 Märstetten - Gugel I).*

Two examples of simulated (blue line) and observed heads (black dots) are shown in Figure 3. In the example shown in the top panel, the head fluctuations are modelled well by the model and the input variables. This is clearly not always the case, as can be observed in the lower panel in Figure 3. For the head time series in the bottom panel, which are affected by pumping, the model performed poorly. Poor model performance can thus be used as an (additional)

indication that other stresses, pumping, river water infiltration, potentially affect the heads at such monitoring stations.

Figure 4 shows a map of the model performances using the coefficient of determination (R^2). The R^2 ranges between 0.13 and 0.96, and a mean and median of 0.61 and 0.64, respectively. Where models have a low performance, it may be used as an indication that one or more other stresses (i.e., pumping, river stage) are impacting the groundwater heads. For these locations, it is unlikely that the heads only show variations due to meteorological changes. An Excel file (notes.xlsx) is provided with an expert judgment on whether it is suspected that stresses other than meteorological stresses likely impact the heads at the stations.

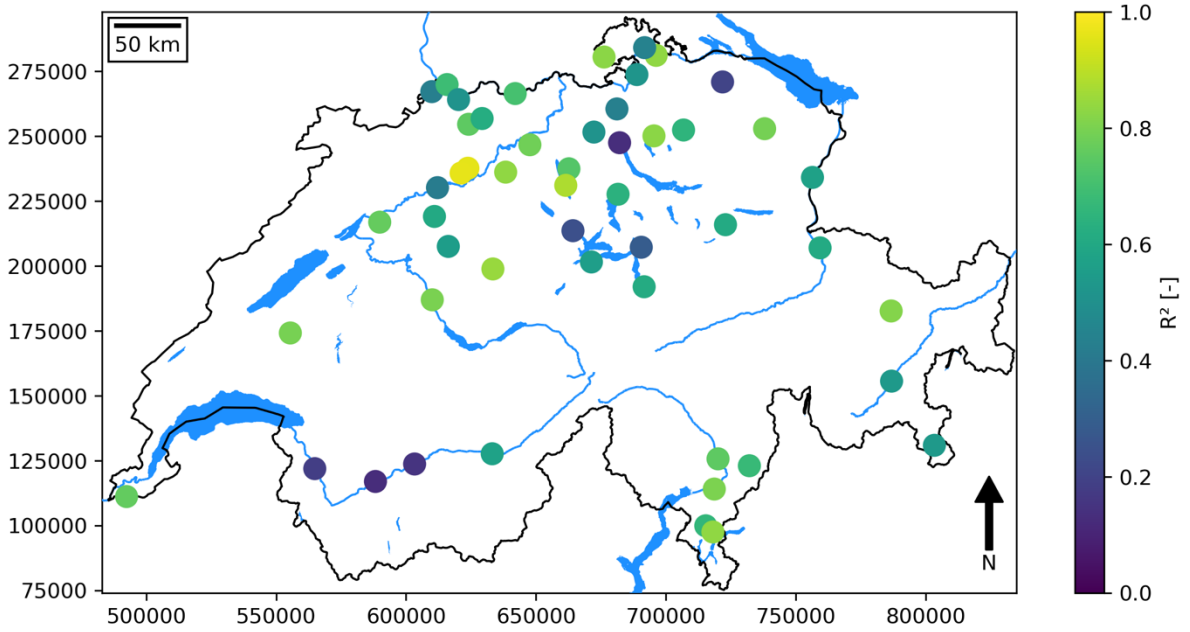


Figure 4. *Map with the model performance quantified using the coefficient of determination (R^2).*

3.2 Gap-filling

The calibrated models are used to gap-fill the 39 head time series that have missing data. Because the simulated heads have a certain bias with the measurement, the simulated data that is used for gap-filling is corrected using the last and first available measurement before and after the gap, respectively. By linearly interpolating the errors at this point, a correction is computed for the simulated heads used to fill the gap. This step was found to be essential for the gap-filling. Figure 5 shows an example of a gap-filled time series for the monitoring station 6511 Davos. The red lines at the bottom mark where gaps were filled. In total, 39 out of the 51 head time series were gap-filled. The final result is a gap-filled time series for all monitoring stations in the dataset.

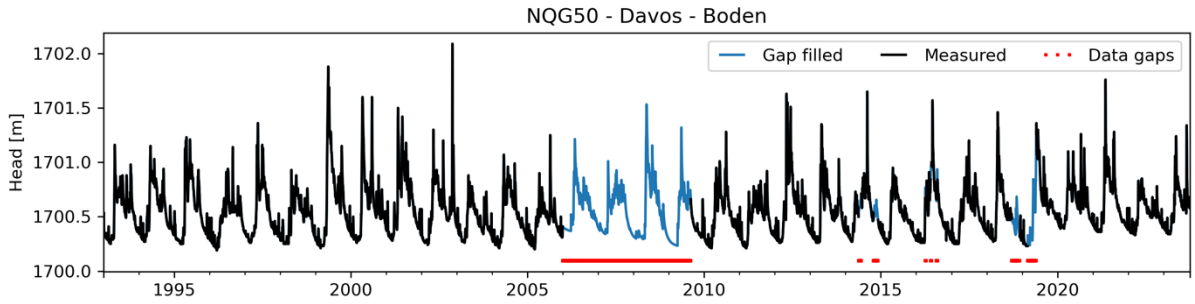


Figure 5. *Example of a gap-filled head time series for the monitoring station in 6511 Davos (NQG50).*

3.3 Script and output data description

The Jupyter notebook *02_gws_modellierung.ipynb* loads the data from the *'processed'* folder, and is used to model the heads using the Python package Pastas and gap fill the time series. The file '*all_timeseries.csv*', saved in the 'output' folder, contains all the time series generated for this exercise. In the folder *'ts'*, all the time series for the individual monitoring stations are stored. The notebook also produces several figures, saved in the *'figures'* folder. Figures of the individual models (simulated and observed heads) are stored in the *'models'* folder, and figures of the gap-filled head time series are stored in the '*gap-filled*' folder.

4 Standardized Groundwater Index (SGI)

4.1 Calculation of the SGI

The standardized groundwater index (SGI) is an internationally used index to compare the severity of drought events from historic records of hydraulic heads (Bloomfield and Marchant, 2013). The index is designed to use 30-year head time series, like many other types of climate analyses. The gap-filled time series are therefore used to compute the SGI for the period 1993-2023. The advantage of the SGI is that it allows to compare extreme events within a time series, and between time series measured at different locations. For each of the gap-filled time series the SGI was computed using the Pastas software. The SGI values typically range between -3 and 3, where low values indicate dry conditions and high levels indicate wet conditions, compared to the same period in other years. Commonly, values around SGI < -2 are used to identify groundwater droughts.

Figure 6 shows an example of the SGI for the monitoring station of 6535 Kestenholz. The top panel shows the gap-filled head time series, with an approximately one-year gap in 2011. During this gap, the Pastas model predicts the lowest hydraulic head. The middle panel shows the SGI as a black line, but also as a shaded color in the background for easier interpretation. For comparison, the bottom panel shows the groundwater drought level determined from the percentile method by the FOEN. The period where the gaps were filled shows some of the lowest hydraulic heads, and thus a low SGI. Missing this event (i.e., no gap-filling) would alter the relative severity of other drought events, making them more extreme. Including this event (i.e., with gap-filling), makes the other low event less extreme and reveals that the most extreme event actually happened in 2011. This example illustrates how data gaps may impact the results of the SGI and why gap-filling can be helpful.

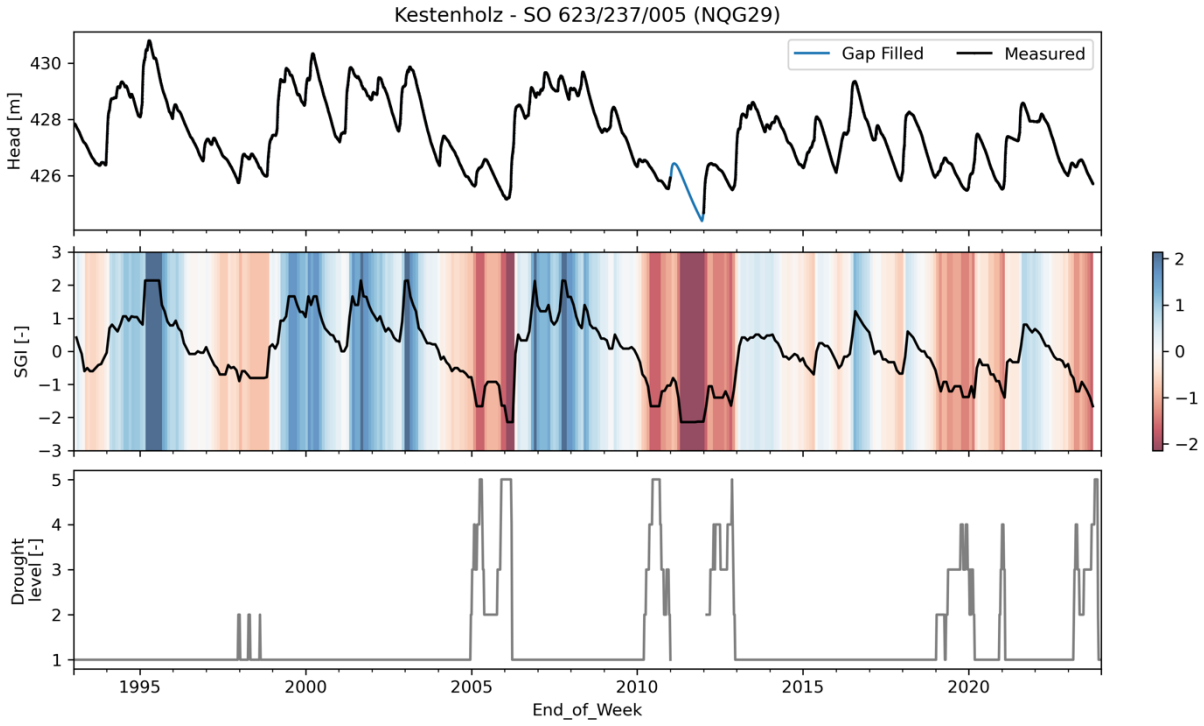


Figure 6. *Time series of the gap-filled heads (top), the SGI (middle), and the groundwater drought level (bottom) for the monitoring station of Kestenholz.*

4.2 Comparison of SGI with percentile-based approach

The FOEN has internally adopted a percentile-based approach to categorize drought events into different groundwater drought levels. With this approach, the head in one particular week of the year is compared to the heads in the same week from a reference period (usually the years 1991-2020). In that sense, the approach is comparable to how the SGI works.

Figure 7 shows a comparison of the SGI plotted against the groundwater drought levels for the analyzed 51 monitoring stations simultaneously. Note that the groundwater drought level has a discrete scale (1 to 5), whereas the SGI is on a continuous scale (~ -3 to 3). It should also be noted that the SGI is computed based on the gap-filled time series, whereas the percentile approach only used the original head data with gaps. Nonetheless, the groundwater drought level and the SGI show a relatively strong correlation (r=-0.61), indicating that both approaches give similar results.

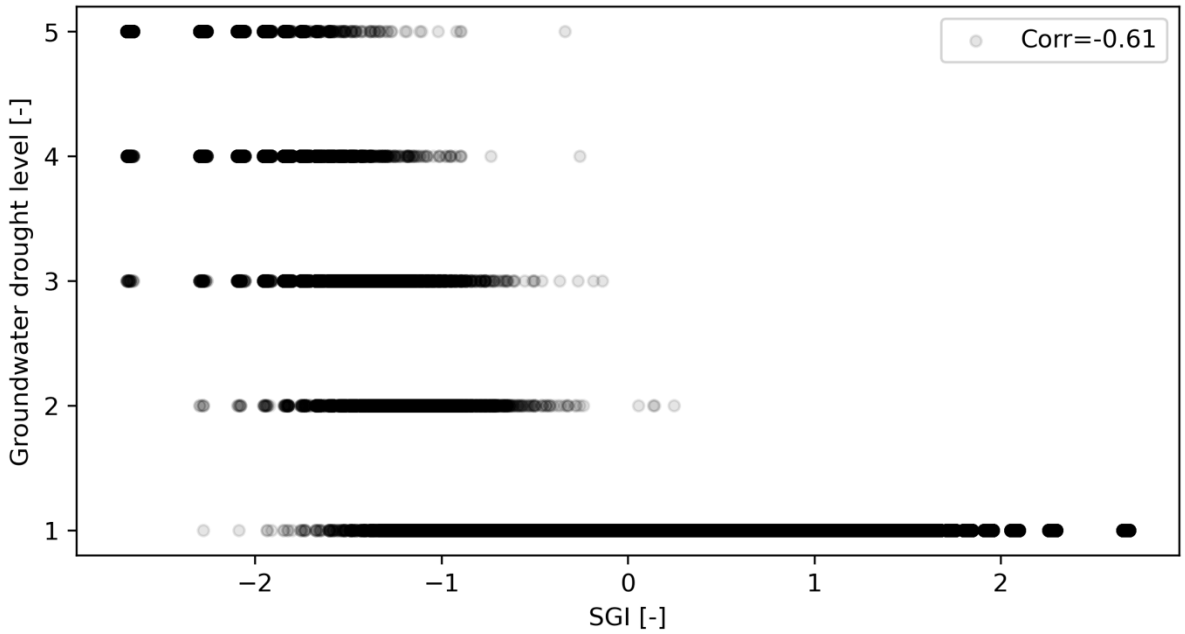


Figure 7. *The percentile-based warning levels plotted against the value of the SGI. The spearman correlation is provided at the top right of the figure.*

4.3 Script and output data description

The Jupyter notebook *02_gws_modellierung.ipynb* that was used in the previous chapter, also computes the SGI and the various figures related to the SGI. The SGI is computed on both weekly and monthly average heads. The resulting SGI values are stored in '*all_timeseries.csv*', saved in the *'output'* folder. The figures of the SGI are stored in the *'sgi'* folder in the *'figures'* folder.

5 Clustering of monitoring stations

5.1 Description of the clustering approach

The goal of the second part of this project was to cluster the different monitoring wells into smaller groups that show similar behavior. For this purpose, the concept of groundwater signatures was used. Groundwater signatures are numerical values that quantify the behavior observed in the head hydrograph (Heudorfer et al., 2019). For example, a signature may quantify whether the head has a long or short memory, or if the head shows a clear annual cycle or not. After computing a set of groundwater signatures, the monitoring stations can be clustered based on the numerical values of the signatures. By selecting different signatures, one can choose what characteristics of the hydrograph to focus on in the clustering exercise.

The approach can be summarized as follows. First, the selected signatures were computed. Second, the signatures were normalized to ensure that the signatures had similar weights for the clustering. Third, a clustering algorithm was applied to cluster the monitoring stations into different groups based on the signatures. The basic principle of such clustering algorithms is to group stations where the signatures have similar values. Here, the Ward method was applied (Ward, 1963), which is a common method for this type of analysis. Here, the monitoring wells were clustered on three characteristics groups: seasonality, response time, and all signatures.

5.2 Seasonality

This clustering focuses on grouping time series with a similar seasonal behavior. For this purpose, signatures focusing on quantifying the seasonal behavior are picked. The three selected signatures are the **average seasonal fluctuation** (avg_seasonal_fluctuation) the **Pardé seasonality** (parde_seasonality) and the **interannual variation** (interannual_variation). The clusters resulting from this analysis are shown in Figure 8, and a map is shown in Figure 9. Figure 8 shows that many of the clusters contain time series with similar patterns. Cluster C0 contains three time series with little to no seasonal pattern. Cluster C1 contains stations with annual cycles combined with much more long-term variations exceeding the annual cycles. These time series show long-term trends that follow the long-term meteorological conditions at the monitoring station. All the stations from cluster C1 are located in northern Switzerland. Clusters C2 and C3 contain time series that appear to respond to individual recharge or precipitation events, but with no clear seasonal pattern. Clusters C4 and C5 on the other hand, show clear annual cycles, likely from recharge events from snow melt in spring and related rises in river levels. The stations in these clusters are all located in the (pre-) Alpine area.

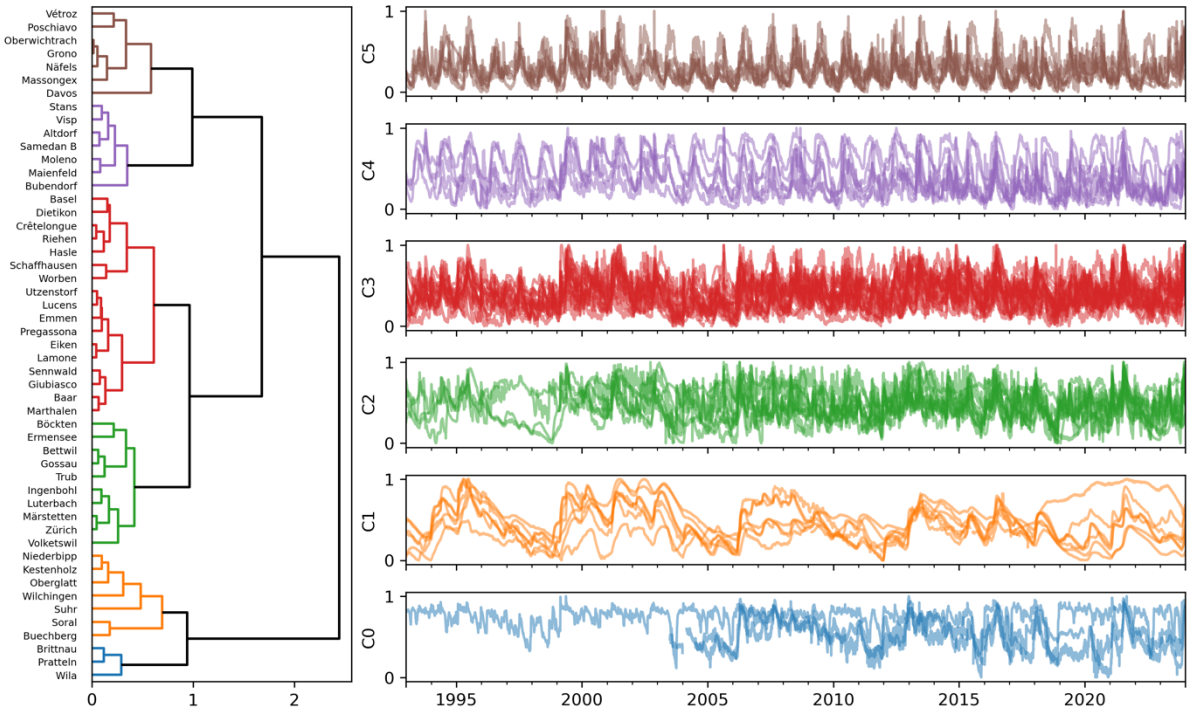


Figure 8. *Clusters based on signatures quantifying the seasonality of the heads.*

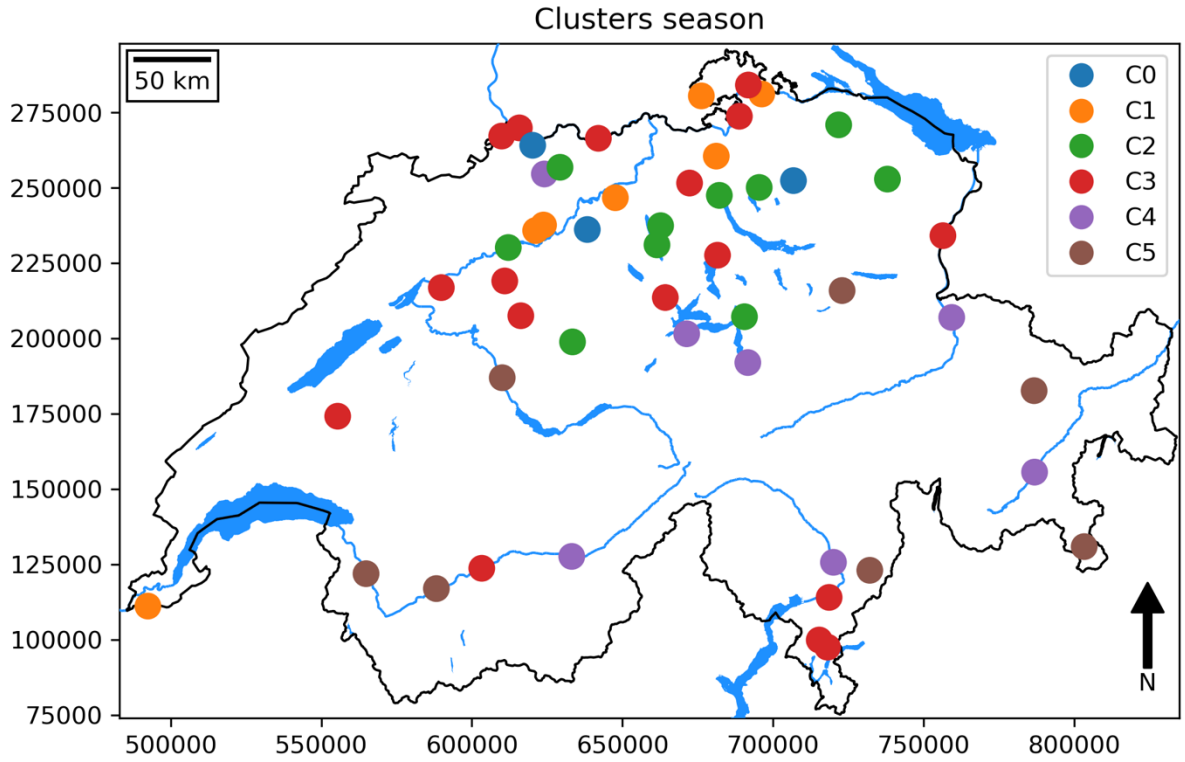


Figure 9. *Map of the clusters based on seasonal signatures.*

5.3 Response time

This clustering grouped the monitoring wells based on their memory or response time, i.e., how quickly does the groundwater system recede and recover? For this, two signatures were

used: the **autocorrelation time** (autocorr_time) and the **fall rate** (fall_rate). The resulting clusters are shown in Figure 10, and a map showing the locations in Figure 11. The hydrographs in each cluster show similar patterns regarding how fast the heads rise and recede. Cluster C3 shows head time series with smooth curves and long memory times, whereas cluster C2 and C1 show decreasing memory times and faster responses, as visible by more peaks. Here, the heads at the stations in clusters C1 respond fastest to recharge events or other stress impulses (i.e., river level changes). The fourth cluster, C0, appears to be a cluster with shorter memory times that did not fit in the clusters C1 and C2. In relation to drought events, it may be expected that the heads in each of these clusters respond faster (C1/C2) or slower (C3) to droughts, and may take shorter or longer, respectively, to recover.

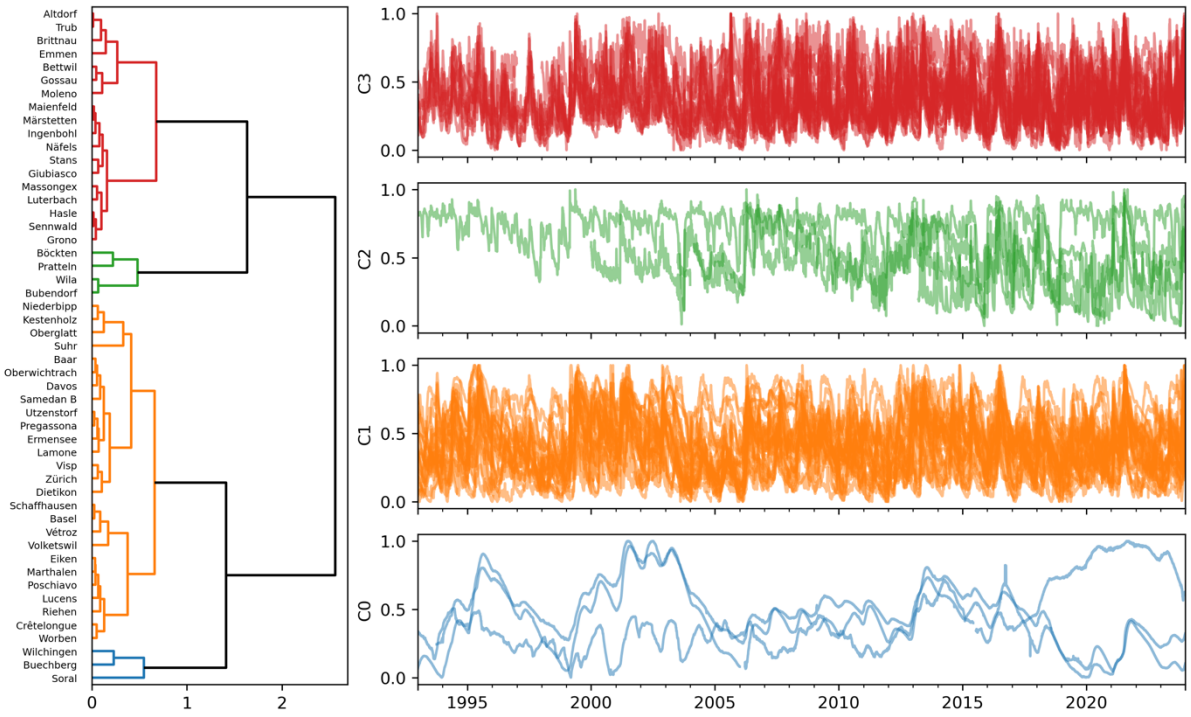


Figure 10. *Clusters based on signatures quantifying the memory time of the groundwater system.*

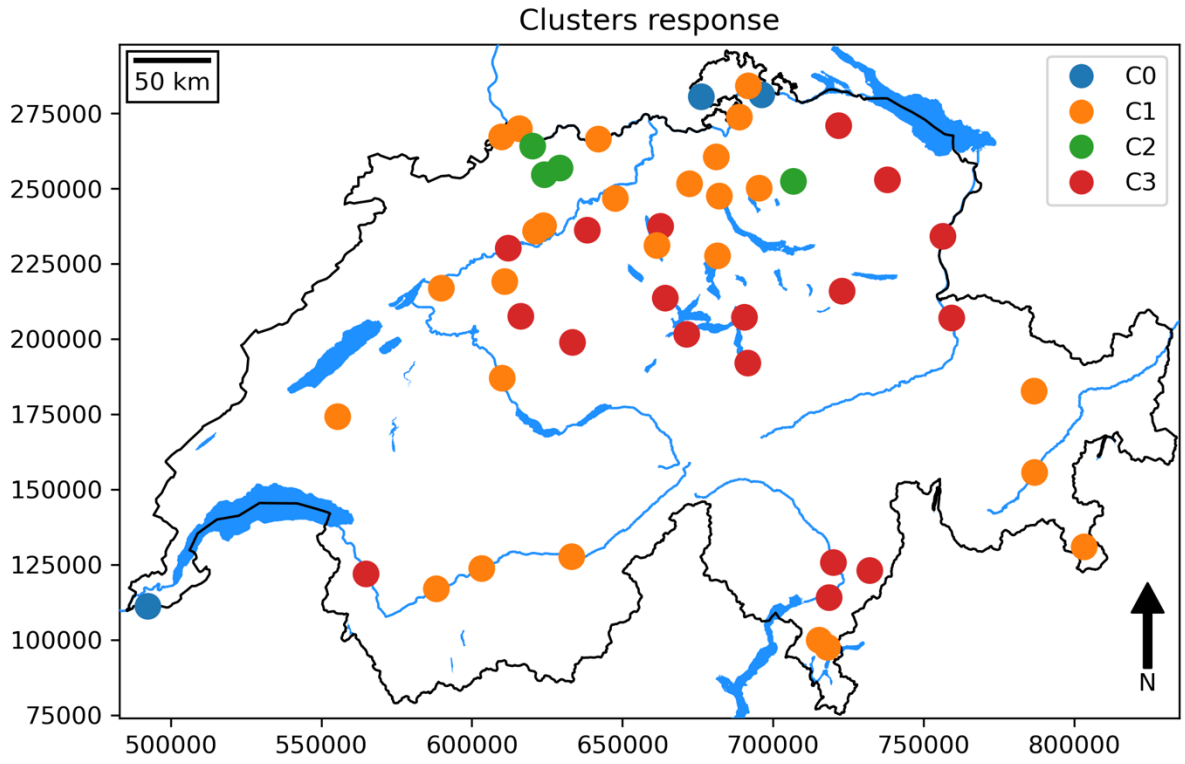


Figure 11. *Map of the clusters based on signatures focusing at memory time.*

5.4 General behavior

For the final clustering exercise, all available signatures were used, except those that showed strong correlations ($r > 0.8$). In the case of strong correlations, one of the two correlating signatures was manually chosen, trying to maximize the total number of signatures by considering correlation with other signatures. This is done to ensure that signatures that focus on similar parts of the hydrograph (and thus are correlated), do not have a higher importance during clustering.

The resulting clusters are shown in Figure 12, and a map of the locations in Figure 13. For this clustering exercise, more clusters were chosen by manually adjusting the number and interpreting the results. Many of the clusters show time series that have very similar behavior, both in terms of memory times and temporal variations. All clusters have two or more stations, except for C9. This cluster represents a station (NQG43) with unique characteristics in terms of the observed head dynamics. The clusters show no clear spatial pattern (Figure 13), but do appear to be related to how far the stations are located away from more mountainous areas. This example of clustering shows that the approach can be employed to cluster stations that observe similar groundwater dynamics, but are not necessarily spatially nearby.

5.5 Script and output data description

The Jupyter Notebook *04_signature_clustering.ipynb* is used for the clustering exercise. This notebook takes the output time series from the previous notebook, computes the signatures, and performs the clustering. The signatures are stored in the file '*signatures.csv*' in the *'output'* folder. The clusters are stored in the '*clusters.csv*' file in the same folder. The notebook

05_create_maps.ipynb is used to produce the maps shown in this report. All figures are stored in the *'figures'* folder.

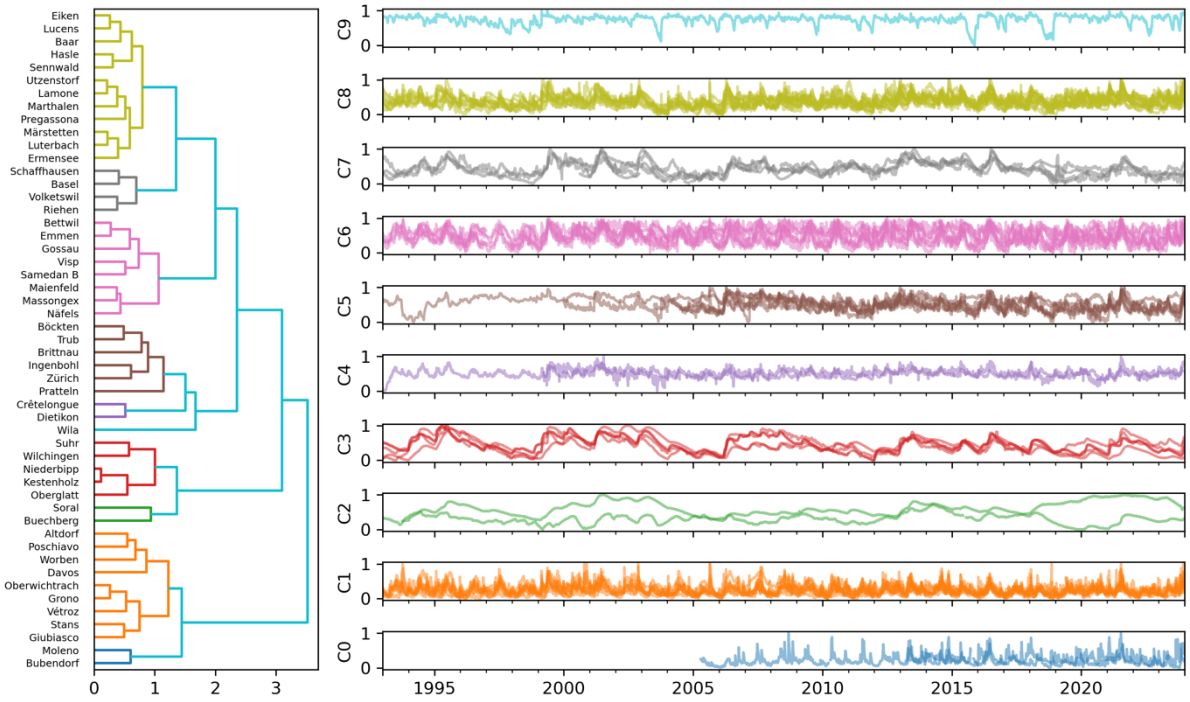


Figure 12. *Clusters based on all signatures excluding those with high correlations between signatures.*

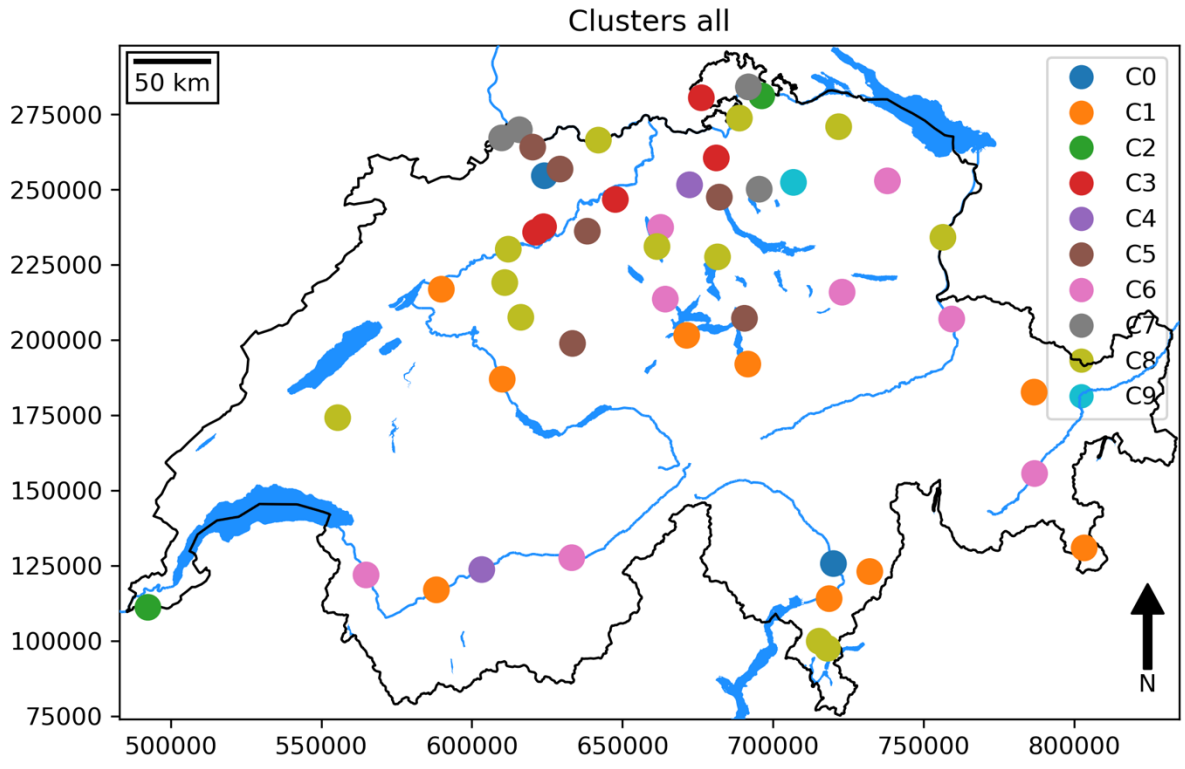


Figure 13. *Map of the clusters based on all the signatures.*

6 Summary

This study investigated the use of the standardized groundwater index and clustering based on groundwater signatures to analyze hydraulic head time series from 51 NAQUA QUANT stations. This report summarizes the main methods and is meant to guide through the data and Python scripts that are the main product of this study.

To compute the standardized groundwater index (SGI), the head time series were gap-filled using lumped-parameter groundwater models and meteorological forcing data. These models additionally provided insights into where stresses besides meteorology might affect the heads. From this analysis, it was found that more than half of the stations measure heads that are affected by other stresses (i.e., river levels or pumping). The gap-filled data was used to compute the SGI, which was compared to the percentile-based approach to quantify a drought level that is internally used within the FOEN. The results of these two approaches were found to be comparable in a qualitative analysis.

In the second part, a novel approach to cluster the monitoring stations, based on the behavior of the head dynamics, was explored and tested on the data. To this end, a large number of groundwater signatures were computed to quantify the behavior observed in the hydraulic head time series. A clustering algorithm was applied to cluster the monitoring station. Based on the selected signatures, clusters could be made on focusing on different parts of the hydrographs. Here, the stations were clustered based on signatures focusing on seasonality, memory time, and the general behavior (many signatures).

Depending on the chosen signatures, different clusters could be identified that group stations where the heads show similar aspects. It is concluded that this type of clustering can be helpful when needing to group stations for a certain behavior, for example the response time to drought events. For a practical application, however, the goal of the clustering needs to be clearly defined, such that signatures can be selected, and the resulting clusters can be evaluated with a specific goal in mind.

7 References

- Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., & Schaars, F. (2019). Pastas: Open source software for the analysis of groundwater time series. *Groundwater*, *57*(6), 877-885.

- MeteoSwiss (2023) RhiresD and TabsD gridded precipitation and temperature data sets. URL: https://www.meteoswiss.admin.ch/climate/the-climate-of-switzerland/spatial-climate-analyses.html

- Vremec, M., Collenteur, R. A., & Birk, S. (2024). PyEt v1. 3.1: a Python package for the estimation of potential evapotranspiration. *Geoscientific Model Development*, *17*(18), 7083-7103.

- Rudolph, M.G., Collenteur, R.A., Kavousi, A. *et al.* (2023) A data-driven approach for modelling Karst spring discharge using transfer function noise models. *Environ Earth Sci* 82, 339. https://doi.org/10.1007/s12665-023-11012-z

- Bloomfield, J. P. and Marchant, B. P. (2013) Analysis of groundwater drought building on the standardised precipitation index approach, Hydrol. Earth Syst. Sci., 17, 4769–4787, https://doi.org/10.5194/hess-17-4769-2013.

- Heudorfer, B., Haaf, E., Stahl, K., & Barthel, R. (2019). Index-based characterization and quantification of groundwater dynamics. *Water Resources Research*, *55*(7), 5575-5592.

- Ward, J. H., Jr. (1963), Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 236–244.