Multicomponent statistical analysis to identify flow and transport processes in a highly-complex environment

Christian Moeck[1], Dirk Radny[1], Paul Borer[1], Judith Rothardt[1], Adrian Auckenthaler[2], Michael Berg[1], Mario Schirmer[1,3]

Contact: Christian.moeck@eawag.ch

[1] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[2] Office of Environmental Protection and Energy, Canton Basel-Country, Switzerland

[3] Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland

Keywords: Hydrochemistry, Groundwater, Multicomponent statistical analysis, Stable water isotopes, Organic micropollutants, Artificial infiltration

This document is the accepted manuscript version of the following article:
Moeck, C., Radny, D., Borer, P., Rothardt, J., Auckenthaler, A., Berg, M., & Schirmer, M. (2016). Multicomponent statistical analysis to identify flow and transport processes in a highly-complex environment. Journal of Hydrology, 542, 437-449. https://doi.org/10.1016/j.jhydrol.2016.09.023

Abstract

A combined approach of multivariate statistical analysis, namely factor analysis (FA) and hierarchical cluster analysis (HCA), interpretation of geochemical processes, stable water isotope data and organic micropollutants enabling to assess spatial patterns of water types was performed for a study area in Switzerland, where drinking water production is close to different potential input pathways for contamination. To avoid drinking water contamination, artificial groundwater recharge with surface water into an aquifer is used to create a hydraulic barrier between potential intake pathways for contamination and drinking water extraction wells. Inter-aquifer mixing in the subsurface is identified, where a high amount of artificial infiltrated surface water is mixed with a lesser amount of water originating from the regional flow pathway in the vicinity of drinking water extraction wells. The spatial distribution of different water types can be estimated and a conceptual system understanding is developed. Results of the multivariate statistical analysis are comparable with gained information from isotopic data and organic micropollutants analyses. The integrated approach using different kinds of observations can be easily transferred to a variety of hydrological settings to synthesise and evaluate large hydrochemical datasets. The combination of additional data with different information content is conceivable and enabled effective interpretation of hydrological processes. Using the applied approach leads to more sound conceptual system understanding acting as the very basis to develop improved water resources management practices in a sustainable way.

1. Introduction

Drinking water supply in urban areas is challenging due to different kinds of water use and potential groundwater contamination and is recognised as a very complex issue including different spatial and temporal scales (Diamond and Hodge, 2007). Historical contaminated sites are often the major concern for a variety of different pollutions in groundwater (Schirmer et al., 2013). These contaminations are problematic especially in the vicinity of drinking water

production sites due to potential risk for human health and the environment. To protect drinking water production sites in an effective and sustainable way, knowledge about flow and transport processes in the heterogeneous subsurface is crucial. It is of vital importance to understand the presence, movement and persistence of contaminations in aquifers to develop adequate groundwater protection plans (Levison et al., 2012). However, studying these processes can be difficult due to their complexity (Atchley et al., 2014; Hunt et al., 2007; Koltermann and Gorelick, 1996; Loague et al., 1998; McCallum et al., 2015; Serrano, 2003, among others) and therefore strategies to protect drinking water production sites are often based on limited data and system knowledge (Mackay et al., 1985).

As a first and basic step for understanding flow and transport processes, a conceptual understanding for the underlying subsurface heterogeneity and the resulting hydraulic flow field has to be developed. Frequently, spatial patterns of water types have to be identified in order to develop a conceptual model for transport and water mixing processes. Hydrochemistry data are used as a tool to evolve such a model (Dilsiz, 2006; Suk and Lee, 1999; Thyne et al., 2004). Mostly, those hydrochemistry patterns have been identified with graphical methods. However, these graphical methods lack clarity when large datasets are considered (Güler et al., 2002; Moya et al., 2015). The interpretation of these graphical methods is typically subjective. To overcome this limitation the use of multivariate statistical analysis, namely factor analysis (FA) and hierarchical cluster analysis (HCA) can be applied. Multivariate statistical analysis has been successfully applied in numerous studies dealing with a wide range of environmental issues (Cloutier et al., 2008; Daughney et al., 2012; Güler et al., 2002; King et al., 2014; Koh et al., 2009; Montcoudiol et al., 2015; Morgenstern et al., 2015; Obeidat et al., 2013; Panda et al., 2006). Cloutier et al. (2008) and Montcoudiol et al. (2015) applied multivariate statistical analysis to investigate the groundwater hydrochemistry of a regional study area to identify the primary processes responsible for groundwater quality and to develop a conceptual model for groundwater flow and geochemical evolution. Daughney et al. (2012) show the application of

multivariate statistical analysis as a means of assessing the representativeness of a groundwater quality monitoring network. A recently published study (Morgenstern et al., 2015) combined multivariate statistical analysis with age dating to understand sources of nutrient contamination for a study site. Hynds et al. (2014) applied HCA to determine clusters based on physico-chemical data to link with the prevalence of bacteria in domestic wells. Selle et al. (2013) studied the spatial and temporal pattern of FA scores to improve the understanding of processes leading to changing groundwater quality in a catchment.

Many past studies have analysed the chemical composition of groundwater to identify hydrochemical processes by applying multivariate statistical analysis in isolation. The validation of the obtained results and adjustment based on other data such as trace components which can have different information content is deficient. Complete 3D information about flow and transport processes are never obtainable (Koltermann and Gorelick, 1996; Moeck et al., 2015) but adequate knowledge about these processes is required for a sustainable water resources management. Remediation strategies as well as mathematical models used to make predictions to evaluate the potential risk of contaminant breakthrough at water supply systems depend heavily on the conceptual process understanding at the investigated site. Once a conceptual model has been constructed according to certain information, it cannot be used to quantify the limitations (Doherty and Simmons, 2013).

Therefore, conceptual process understanding should not rely on a limited data set only, but should rather take a combined approach of different sets of data which are independent to each other to produce the most credible description of flow and transport processes.

Thus, in this study we use a combined approach of multivariate statistical analysis, interpretation of geochemical processes, stable water isotope and organic micropollutants to develop a credible conceptual process understanding, which should be a basis for further field surveys and numerical modelling experiments. Since the composition of stable water isotopes is not altered in the aquifer by water-rock interactions, this provides us with an additional tool

to understand mechanisms of groundwater circulation such as mixing between different aquifer systems (Chen et al., 2011; Demlie et al., 2007). Furthermore, organic micropollutants were used to estimate specific water pathways (Huntscha et al., 2013; Schirmer et al., 2011; Van Stempvoort et al., 2011) and were applied as a tracer to identify spatial distribution and spreading of different water types (Scheurer et al., 2011).

The objective of this study was to compare and combine obtained information from different data types and sources to develop a consistent conceptual process understanding. The benefit of having such a high-quality and quantity data set is demonstrated for a study area in Switzerland, where drinking water production is close to different former landfills and industrial areas. The study site is characterized by a highly-complex geological and tectonical setting with a partly-connected gravel and karst aquifer. To avoid drinking water contamination, artificial groundwater recharge with surface water into the gravel aquifer is used to create a hydraulic barrier between contaminated sites and drinking water extraction wells. Nevertheless, a range of organic micropollutants, which nowadays lack in the surface water, can be measured in the abstracted groundwater. Therefore, it is important to understand the subsurface heterogeneity and processes to ensure the utmost security for drinking water production although most of the micropollutents detected are below drinking water thresholds in Switzerland. More specifically, the following research questions and tasks are addressed:

I. The spatial distribution of hydrochemical data is systematically assessed based on a multivariate statistical analysis.
II. How geochemical processes can be linked to the spatial distribution of hydrochemistry data and present geological information is assessed to evaluate the influence of groundwater recharge, inter-aquifer mixing and fault induced aquifer connectivity along the flow paths.
III. The obtained results of the multivariate statistical analysis are compared to the observed concentration patterns of stable isotopes and organic micropollutants

IV. The origin of water at the drinking water pumping gallery which will ensure the utmost security for drinking water under actual water management strategies is determined.

2. Study Area

The study area *Hardwald* is located in north-western Switzerland with an area of approximately 10 km^2 (Fig.1). Two main aquifers exist namely the Quaternary Rhine gravel and the underlying karstified Upper Muschelkalk limestone aquifer (see section 2.1). The Quaternary gravel aquifer is primarily used for drinking water production combined with an artificial groundwater recharge system. The infiltration system was designed in order to maintain a hydraulic gradient towards areas of potential risk which are highlighted in Figure 1. Furthermore, large quantities of groundwater are abstracted for industrial purpose east of the study site.

Figure 1: Top panel: Study area in north-west Switzerland with land use, groundwater sampling locations and Rhine and Birs rivers. Elevations range from south to north from around 290 meters above sea level (masl) to 250 masl in the vicinity of the Rhine. The northern boundary of the Hardwald is the river Rhine, urban areas and river Birs in the west, railways and industrial areas in the south as well as industrial areas in the east. A highway passes directly through the Hardwald from west to east. Yellow circles show piezometers, red triangles extraction wells. Urban settlements surrounding the study site in the west and the south are illustrated as grey shaded areas. Artificial infiltration channels and ponds are shown in light-blue. The piezometric heads are shown as black dashed lines. Bottom panel: Simplified geological bedrock map of the study area including the rivers Rhine and Birs. Artificial

infiltration channels and ponds are shown in light-blue. A simplified cross section is shown in the supporting information (Figure S1).

2.1 Geological Setting of the Study Area

The geological structure in the study area can be subdivided into different components from west to east (Fig. 1 lower panel and Figure S1, supporting information). The western part is situated to the south eastern shoulder of the Upper Rhine Graben, the so called Rhine Valley flexure zone (Fig. 1), an intense structurally deformed area. Further east the Hardgraben is located. The Rhine Valley flexure zone as well as the occurring Horst and Graben structures show a NE-SW orientation (Gürler et al., 1987).

The bedrock are assigned to Triassic strata, which dip slightly to the southeast. The Middle Muschelkalk is subdivided into the Sulfate and Dolomite zone (Spottke et al., 2005). The Sulfate zone is considered to be impermeable (Gürler et al., 1987; Nagra, 2002) although a gypsum karst unit between a salt and an anhydrite layer was observed and might transport water into the salt layer (Aegerter and Bosshard, 1999). The Upper Muschelkalk, consisting mainly of limestone with a thickness of 50-60 m. The Dolomite zone and the limestone of the Muschelkalk represent an important highly fractured and karstified aquifer with a mean hydraulic conductivity of $1.3*10^{-4}$ m/s (Gürler et al., 1987, Spottke et al., 2005) but within a range between $1*10^{-3}$ m/s and $2*10^{-6}$ m/s (Affolter et al., 2010; Huggenberger et al., 2009) depending on the location.

The overlying Jurassic strata are made of 20-40 m thick porous limestone (Lias) and of an 80-100 m thick clay layer (Opalinuston). These units represent an aquiclude due to their low hydraulic conductivity between $1.3*10^{-14}$ to $1.3*10^{-7}$ m/s (Nagra, 2002). In the study area, the Jurassic strata is present only in the Hardgraben whilst eroded in the surrounding study area.

The fluvio-glacial deposits, mainly gravels from Quaternary strata appear with a varying thickness of 5 to 50 m on top of the bedrock. The thickness close to the main drinking water extraction wells (i. e. the pumping well gallery, Fig. 1) is around 20 to 40 m. The average hydraulic conductivity is $3.1*10^{-3}$ m/s (Spottke et al., 2005). A funnel shaped depression of the surface can be found in the western part of the study area. These features are interpreted as karst sinkholes replenished by gravel deposits (MBN, 2008).

The hydraulic conductivity of the aforementioned fracture zones is unknown and might be variable spatially. It may act as an additional flow path due to increasing permeability and therefore support vertical exchange of groundwater although an aquitard exist (Gürler et al., 1987) or it is filled with fine sediments and builds a hydraulic barrier. Furthermore, the Sulfate zone with its low hydraulic conductivity can lead to upwelling of for instance Muschelkalk groundwater into the sand gravel aquifer (Spottke et al., 2005). Using a 3D groundwater flow model in the same study area Spottke et al. (2005) speculated that the abstracted water used by large-scale industrial pumping in the western part of the study area must contain water fractions from the deeper Muschelkalk aquifer. They noticed that the productivity of industrial pumping wells can only be supported by groundwater recharge from adjacent horsts inducing a circular flow pattern or upwelling. Therefore, knowledge of the connectivity and interaction between different aquifer systems are fundamental for understanding the ongoing processes in the study area. A more detailed description of the geology and tectonics in the study area and its surrounding is given in Spottke et al. (2005).

2.2 Hydrogeology and Historical Context of the Study Area

The highly-complex environment *Hardwald* in terms of artificial infiltration, pumping water abstraction for both, drinking water supply and industrial proposes combined with a complex geology has changing flow boundary conditions in the last decades. The natural groundwater flow is influenced by two changing boundary conditions. Until 1958 the natural groundwater

flow was from south to north in the direction of the river Rhine (Figure S2a, supporting information). The other river Birs is located in the western part of the study area. The influence on the flow field at the study area is likely low due to the distance between *Hardwald* and Birs river. Due to the installation of a hydropower plant downstream of the study area, the flow direction changed. The hereafter established flow was altering from southeast towards northwest (Figure S2b, supporting information).

Increasing water demand of the growing population and industry led to the installation of an artificial groundwater recharge system in 1958. Here, Rhine water is pumped to an excavated system of channels and seven infiltration ponds to replenish the aquifer (Fig 1). Currently, the infiltrated volume averages 100,000 m^3/d. Apart from water quantity issues, the quality is also influenced by the recharge amount. Due to the immediate vicinity to former landfills and industrial sites, this artificial recharge is used to minimise the risk of drinking water contamination. This is accomplished by recharging twice as much water as abstracted. This approach of artificial infiltration leads to an elevated local groundwater mound, which serves as a barrier preventing natural inflow of potentially contaminated water coming from adjacent areas (Figure S2c) (Auckenthaler et al., 2010). However, trace levels of some contaminants can be found in the groundwater. Note that all concentrations of detected contaminants at the pumping wells are very low and mostly far below the limits of relevant drinking water regulations. In addition, activated carbon treatment of the abstracted water ensures that contaminants are removed to the most possible extent (Gabriel and Meier, 2014). To provide furthermore water for industrial use a pumping well south of the *Hardwald*, beneath a landfill is operated since 1957. This industrial water use avoids to a great extent the spreading of contaminated water from the landfill nearby.

Different potential sources of contamination in the extracted water were taken into consideration. A common hypothesis is that the present observed concentrations of

Tetrachloroethylene (PCE) and Trichloroethylene (TCE) were released from adjacent former landfills south from the pumping water gallery (Auckenthaler et al., 2010). The bedrock at these former landfills consists mainly of limestone from the fractured Muschelkalk formation, which is the basis of the landfills. According to Levison and Novakowski (2012) fractured bedrock aquifers with minimum or without overburden cover are highly vulnerable to contamination. It is assumed that the VOCs (Volatile Organic Compounds) were transported towards the river Rhine before artificial infiltration of surface water was established and contaminated the aquifer down to a depth of 100 m where they still can be found in groundwater samples (Auckenthaler et al., 2010).

A hypothesis for the origin of Tetrachlorbutadiene (TeCBD) and Hexachlorbutadiene (HeCBD) in the groundwater is that they are remobilized from the quaternary Rhine sediments. Until the 1980s, the Rhine water used for artificial groundwater recharge was not as clean as nowadays and present contamination patterns might be related to remobilisation effects. Stieglitz et al. (1976) reported 4.58 µg/l of 1,1,4,4 Tetrachlorbutadiene isomer (TeCBD) in Rhine water of the mid-1970s. This TeCBD concentration was higher than TeCBD concentrations ever found during our investigation in the study area. The current highest measured concentration in groundwater is 0.22 µg/l.

The study focusses on the contaminants which are nowadays most frequently found at the groundwater extraction wells. These are i) Tetrachloroethylene (PCE) and Trichloroethylene (TCE), two of the most widely used cleaning and degreasing solvents in Switzerland and in many parts of the world (Doherty, 2000) and ii) Hexachlorbutadiene (HeCBD) and 1,1,4,4 Tetrachlorbutadiene (TeCBD), by-products of the manufacture of chlorinated solvents (Fields, 2004). The isomer 1,1,4,4 TeCBD can further be produced as a result of HeCBD dechlorination (Bosma et al., 1994). All these compounds have a higher density than water ($\rho > \rho_{water}$) and percolation into the subsurface can occur. Further information on the above mentioned

components can be found in the listed references (Avon and Bredehoeft, 1989; Bosma et al., 1994; Fields, 2004; Gandhi et al., 2002a; Gandhi et al., 2002b; Mayer et al., 2001; Messier et al., 2012; Murphy et al., 2011; Ninomiya et al., 1994; Poulsen and Kueper, 1992, among others).

3. Methodology

The location of the drinking water production area in the direct vicinity of former landfills and industrial sites has resulted in an extensive monitoring of the groundwater over several years and therefore a large amount of physico-chemical and chemical data is available. This large amount of data is unfortunately spatially and temporally inconsistent and makes a grouping and comparison of groundwater samples challenging. To incorporate a systematic investigation of the different data, a multivariate statistical analysis was chosen. Here, two well established multivariate methods were applied using the open source software 'R' (http://www.r-project.org/): The factor analysis (FA) and the hierarchical cluster analysis (HCA). The results of these analyses were combined with gained information from stable water isotope and organic micropollutant analyses. This combination was furthermore used to identify processes controlling the groundwater quality in the study area. Overall, the relation between the groundwater chemistry and the spatial distribution of selected contaminants was assessed and its robustness further validated using the t-test. In the following section the hydrochemistry for the study area is described as well as the required data preparation for the FA and the HCA. Subsequently, the methodology of using FA and HCA is provided.

3.1 Datasets and Sampling

The sampling locations in the *Hardwald* are distributed over the whole study area and comprise 43 piezometers, 6 multi-level wells, 34 extraction wells and 2 rivers namely Rhine and Birs (Figure 1). In total, 85 sampling locations were considered. The piezometer and well depths varies between 17 (absolute altitude 255 masl) and 80 m (absolute altitude 188 masl) and are

drilled mostly until the bedrock was reached (see Table S1, supporting information; 25 sampling locations are screened in Quaternary formations, 58 wells are drilled in the bedrock and one is unknown).

The statistically analysed hydrochemical data originates from two different sources. The first group of data was provided from the cantonal authorities. This group includes hydrochemistry data from 2005 to 2013 that covers 85 sampling locations.

Additionally to these data, we conducted two sampling campaigns in the *Hardwald* in November 2014 and March 2015 with 37 and 43 groundwater sampling locations, respectively. The samples were collected (1) by using stainless steel submersible pumps (Grundfos, MP1), (2) directly from the tap in the well houses of the extraction wells or (3) as grab samples in the rivers. In the field, the physico-chemical parameters temperature, pH, electrical conductivity, and the oxidation-reduction-potential (ORP) as well as dissolved oxygen concentrations were measured using appropriate devices (HACH-LANGE, HQ40d). Sample collection was conducted after stabilization of the physico-chemical parameters. The water samples were analysed for major cations- and anions as well as trace elements and organic micropollutants (full list of organic micropollutants are given in Table S2, supporting information). In addition, samples for stable isotope (deuterium ($\delta\ ^{2}H$) and oxygen ($\delta\ ^{18}O$)) analysis were taken and analysed. The chemical analysis was performed using standard methods and devices for the major cat- and anions (Metrohm 761 Compact IC), the trace elements (Agilent ICP-MS 7500 cx), organic micropollutants (Q-Exactive Plus) and the stable isotopes (Picarro L1102-I). All trace elements, stable isotopes, cat- and anions samples acquired during the two field campaigns were analysed at EAWAG labs. Triplicate measurements and blanks were used, but the results were very stable and no data drift was observed. The provided data from the cantonal authorities were analysed in the accredited cantonal lab

3.2 Data Preparation

Only datasets with a charge balance error under 5% were used for the hydrochemistry analysis and interpretation (Note the majority of samples had a charge balance error of 1%). Güler et al. (2002) and van den Brink et al. (2007) stated that a calculated charge balance under 5% is an acceptable error. In case of multiple analyses for the same sampling location, the mean value of the analyses of a certain parameter/compound was calculated. This procedure is valid as temporal effects are negligible in the study area in comparison to the impact on the spatial variability in hydrochemistry data (see Table S3, supporting information).

Cloutier et al. (2008) pointed out that a certain number of parameters should be excluded in the multicomponent statistical analysis for the following reasons: Parameters with additive characteristics such as TDS (Total Dissolved Solids) and EC (Electrical Conductivity), parameters with an elevated number of samples below the detection limit and most trace constituents and parameters missing or not analysed for a large number of sampling sites. Therefore, the resulting multivariate method was applied on a robust subgroup of the groundwater hydrochemical dataset (without the two rivers), including 83 sampling locations and 12 parameters. The selected parameters in this study comprise Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, HCO_3^{-}, Cl^{-}, NO_3^{-}, SO_4^{2-} as well as the trace elements Sr^{2+}, As^{3+}, U^{6+}, and Li^{-}.

The descriptive statistics of the data matrix is given in Table S4 (supporting information). Most parameters are positive skewed and therefore a log transformation was required (Güler et al., 2002; Montcoudiol et al., 2015). Subsequently all parameters were scaled to obtain new values which have zero mean and measured in standard deviation units (Davis, 1986). Using this procedure all chemical parameters have equal weights for the multivariate statistical analysis to account for the different order of magnitude between major ions (mg/l) and trace elements (µg/l).

The processed standardized dataset was used in a first step for the factor analysis (FA) to deduce interrelationships between parameters. For the FA the commonly applied rotation method "varimax" was used, which seeks to maximize the variance of the squared loading for each factor. The calculated eigenvalues and eigenvectors illustrate which parameters are showing the same trend. The choice of the total amount of factors to describe the variance of the dataset is uncertain. Typically factors with an eigenvalue greater than one are retained. These factors explain most of the variance and eigenvalues and can be considered as the dominate parameters for differences in the hydrochemistry. Those less important were discarded (Davis, 1986). For our dataset only the first three factors were further investigated since they account for more than two-thirds of the explained variance.

In a second step the standardized dataset was used to estimate groups of samples with similarities using a linkage rule for the HCA. There is a wide range of HCA approaches but a common methodology to obtain distance measures with linkage rules is the Euclidian distance together with the Ward's method (Daughney et al., 2012; Montcoudiol et al., 2015; Panda et al., 2006; Singh et al., 2008). Acording to Güler et al. (2002) this approach gives the most distinctive groups. The resulting dendrogram displays the linkage between the samples according to their similarities (Davis 1986). We determined the number of clusters based on the proposed approach of Hothorn and Everitt (2009) where the number of clusters are chosen and estimated on a graphical decision. In order to ensure a reliable dendogram we further applied multiscale bootstrap resampling (Suzuki and Shimodaira, 2006). This multiscale bootstrap resampling implemented in the R software environment helps for assessing uncertainty in hierarchical cluster analysis. Calculated probability values (p-values) between 0 and 1 for each cluster indicate finally how strong the cluster is supported by the available data. Only cluster which seem to be reliable were used for further analysis and interpretation.

4. Results and Discussion

In the following section observed concentration patterns and ranges for chosen contaminations are described. Then the results of the multivariate statistical analysis, namely FA and HCA are discussed. These results are linked to hydrochemical processes in the subsurface, such as dilution and mixing, using patterns of major ions as well as stable isotope and organic micropollutants data. Subsequently obtained information from the aforementioned data source was related to the observed contamination patterns.

4.1 Patterns

The observed concentration range for TCE and PCE in the groundwater is between 0.14 to 20 µg/l and 0.14 to 54 µg/l, respectively (Figure 2). The highest concentrations can be found south and at the western edge of the study area. The lowest concentrations can be found between the artificial infiltration system and the extraction wells with values smaller than 0.30 µg/l. A quite similar distribution can be found for HeCBD. Highest concentrations are located south with around 3 µg/l whereas the concentration tends to 0.01 µg/l between the artificial infiltration system and the extraction wells. Slightly higher concentrations appear at the western edge of the study area, especially in the northern part. The spatial distribution of TeCBD shows visually the same trend as observed for HeCBD but with smaller concentration ranges between 0.01 and 0.14 µg/l. The only contrast in the spatial distribution of HeCBD and TeCBD is noticed in the south-west, where low concentrations of TeCBD can be found as opposed to higher concentrations of HeCBD. Note that a smaller number of analyses exist for TCE (N=71) compared to the remaining compounds (N=82-83), especially at the western edge of the study area. Independent of their spatial distribution, at the Multi-level wells we can observe that with increasing depth the concentrations of all contaminants increase, although the differences are relatively small (Figure S3-S6, supporting information).

*Figure 2: Spatial linear interpolation of (log) concentration (*µg/l*) for PCE, TCE, HeCBD, and TeCBD. In case of multiple analyses for the same sampling location, the mean value was used. Red colours indicate higher concentrations, whereas blue colours lower concentrations. Note that the colour scale is different for all compounds. Gray lines show the concentration contours. The solid thick white lines indicate the location of the infiltration channels and ponds, whereas the blue solid line shows the position of the river Rhine.*

Correlations were observed between PCE, TCE, HeCBD and TeCBD for all groundwater sampling locations (Figure S6, supporting information). A high correlation can be found for PCE and TCE with a value of 0.84. Also PCE shows a good correlation with HeCBD with 0.67. Differences in the correlation might be attributed to different properties of these compounds, which can lead to slightly different concentration patterns and therefore smaller correlation. The strongest correlation can be found for HeCBD and TeCBD of 0.88. This strong correlation might indicate degradation from HeCBD to TeCBD as demonstrated by Bosma et al. (1994) or indicate similar transport processes. However, which process leads to the observed concentrations cannot be fully clarified with our data and analysis.

4.2 Multivariate Statistical Analysis

In Table 1 the first three factors are shown describing more than two-thirds of the explained variance. The first factor is characterized by high load of Na^+, Cl^- and Li^+ whereas factor two shows a high load in Ca^{2+}, HCO_3^- and NO_3^-. Factor three, with a smaller contribution to the variance indicates a high load in Sr^{2+} and SO_4^{2-}.

Table 1: First three factors with associated load of the cations and anions of the factor analysis. These three factors describe 69.2% of the data set variance.

Element	***Factor1***	***Factor2***	***Factor3***
Na^+	0.96	0.17	0.20
K^+	0.56	0.00	0.46
Mg^{2+}	0.36	0.68	0.33
Ca^{2+}	0.21	0.94	0.22
HCO_3^-	0.32	0.86	0.06
Cl^-	0.85	0.35	0.27
NO_3^-	0.00	0.78	-0.12
SO_4^{2-}	0.43	0.49	0.67
Li^-	0.81	0.20	0.32
Sr^{2+}	0.34	0.02	0.94
As^{3+}	-0.17	-0.08	0.02
U^{6+}	-0.11	-0.16	0.02
Explained variance	3.2	2.7	2.4
Explained variance (%)	26.9	22.4	19.9
Cumulative variance (%)	26.9	49.3	69.2

Figure S8 (supporting information) illustrates the position of the factor loads of the chemical parameters defined by the axes of factor one and two. Because of the high load of Na^+, Li^+, Cl^- in factor one; the y-axis is defined as the salinity in reference to NaCl. Factor two with high

loads of Ca^{2+},Mg^{2+} and HCO_3^- can be referred to the hardness term. The NaCl distribution shows the spreading of artificial infiltration and changes in the geological units (Figure S9, supporting information). Low values can be observed where the fraction of artificial infiltration with Rhine water is high, whereas higher values occur at the borders of the study area. In these areas it is likely that the water is older than the young infiltrated Rhine water in the vicinity of the water extraction well gallery. The high load of $NO3^-$ in factor two show the influence of the differences between the geology, especially in the northern part. In the northwest the Sulfate zone is located and can be identified with low NO_3^- values (Figure S9, supporting information). The high load of Ca^{2+} and HCO_3^- in factor two is likely related to the Muschelkalk unit in the southern part of the study area (Figure S9, supporting information). Processes leading to their loads are discussed in the following sections.

The results of the HCA are displayed as a dendrogram (Figure 3a) and reveal some level of similarity between clusters. Six different groups were defined based on the so-called phenon line. Cluster 1 to 3 (C1-C3) are quite different from clusters 4 to 6 (C4-C6) due to a higher linkage distance, indicating a distinct different hydrochemical composition. A relatively small linkage distance exists between C2 and C3 as well as between C4 and C6. Note that C4 is just one sample location situated in the vicinity of the Rhine Valley Flexure zone (Fig. 1 and 4), and must be interpreted carefully. This cluster differs compared to the other clusters in terms of high concentrations of Na^+,K^+, Ca^{2+}, Mg^{2+}, HCO_3^-, SO_4^{2-} and Cl^-. It might be an indication of vertical exchange and upwelling of waters originating from the salt and gypsum layer (see Section 2.1). Calculated bicarbonate equilibrium using PHREEQC (Parkhurst and Appelo, 1999) for all samples for the November sampling campaign shows again the well-defined separated cluster group, but also indicate that C4 has a distinct different Ca^{2+}- CO_2 pressure relationship compared to all other clusters which further validate C4 as an exception (Figure S10, supporting information). Furthermore, C4 shows the highest standard deviation based on the temporal

analyses of major cations and anions (Table S3, supporting information). We assume that the vertical exchange might just be temporally activated, depending on the hydraulic conditions within the fracture network and variable artificial infiltration rates at the western channels and ponds.

To describe the mean hydrochemistry of the clusters, Stiff diagrams were used (Figure 3b). The Stiff diagrams show the major ion concentrations (as milli-equivalents per litre) and differences, similarities of main cat- and anions can be withdrawn from the clusters. Thus, the hydrochemistry of C1 and C2 is quite similar to that of the Rhine river water. Clusters 1 and 2 represent groundwater from the Quaternary aquifer, that is: Artificially infiltrated Rhine water in the Quaternary aquifer. In contrast, clusters C4-C6 show higher concentrations in major ions compared to C1-C2, and the Rhine river water, especially with regard to Ca^{2+} and HCO_3^-. Therefore, C6 most likely represents water originating from the Muschelkalk and C4 and C5 represent a mixed groundwater. Also the hydrochemistry of C3 could potentially indicate a mixing between "Quaternary water" from C1-C2 and to a lesser extent Muschelkalk water from C6, and will be further discussed in the following sections. The hydrochemistry of the Birs river seems to be not correlated with all estimated cluster groups.

Figure 3:a) Dendrogram of the HCA for 83 groundwater sampling locations with six different clusters (C1-C6) based on the linkage distance (for clarity, sample names are not displayed). Note that the two rivers were not included. b) Stiff diagram of major ions, concentration as milli-equivalents per litre, for the estimated six different clusters and the two rivers Birs and Rhine.

The spatial distribution of the six clusters shows a distinct trend (Figure 4). C1 and C2 are located close to the artificial infiltration. C3 can be found mainly at the northern west edge of the study area where the effect of the artificial infiltration is decreasing and partly within the pumping gallery. C6 is found in the south close to a former landfill (Muschelkalk limestone

area), and clusters C4 and C5 at the western edge most likely follow the regional flow path from C6.

The origin of water at the pumping gallery is of most interest related to drinking water production. Most extraction wells belong to Cluster C1 and C2 (Artificial infiltrated Rhine water from the Quaternary aquifer). Only four wells, belonging to cluster 3, show a different hydrochemistry compared to C1 and C2 (Figure 4 and Figure S11, supporting information) and as shown in Section 4.6 can be linked to slightly higher observed concentrations of e.g. PCE compared to concentration values for C1 and C2 .

Figure 4: Spatial distribution of clusters according to the HCA. The green circles indicate the four extraction wells that belong to Cluster C3.

4.3 Interpretation of geochemical processes

Enrichment in Ca^{2+} and HCO_3^- can be observed from cluster C1 to C6 (Figure 5). It is assumed that Ca-rich mineral dissolution is the primary cause leading to the observed trend. Due to the present Muschelkalk limestone in the study area, calcite dissolution is the most likely process. This assumption is further supported by assigning the theoretical HCO_3^-:Ca^{2+} line in Figure 5, where data have a slope close to 2:1. Furthermore, C1 and C2 show a quite similar trend as observed for the Rhine river water. Again a strong influence of artificial infiltration on the hydrochemistry can be observed. Here once more, the hydrochemistry of C5 could indicate a mixing between “Quaternary water” from C1-C2 and Muschelkalk water from C6. To a much lesser extent this also could potentially be a process for the observed pattern in C3. Similar well defined patterns as observed for Ca^{2+} and HCO_3^- can be identified also for the Ca^+ and SO_4^{2-} relationship (Figure S12, supporting information)

Figure 5: Plot of HCO_3^- and Ca^{2+} [mmol/l]. The Birs and Rhine surface water concentrations are displayed separately with yellow and blue colours, respectively.

4.4 Validation of the multicomponent statistical analysis with isotope data

Stable water isotopes ($\delta^{18}O$ and $\delta^{2}H$) ranged from -11.33 to -8.68‰ and -81.52 to -63.97‰, respectively (Figure 6a). The isotope compositions follow the local meteoric water line (LMWL: straight dashed line in blue). The water isotopes reproduced again the well-defined cluster groups derived from HCA. C2 and most of the sampling points of C3 shows a strong link with the Rhine river water whereas cluster C5 and C6 have an enriched isotope composition, indicating different water origins and possible ages. Cluster C5 and partly C3 represent most likely mixtures of at least two water types, which could be C6 and C2. Although the water isotopes from the Birs river show values between the clusters group C1-3 to C5-6, it is unlikely that the Birs river has a strong impact on the flow field or hydrochemistry of the study area. We observed already that the hydrochemistry in cat- and anions is different and not comparable to any other sampling location.

The interrelationship between $\delta^{18}O$- HCO_3^- shows a high correlation of 0.84 (Figure 6b). High values for the HCO_3^- can be found for C6 associated with relatively enriched values for $\delta^{18}O$ whereas the low concentration of HCO_3^- can be found for C2 related with relatively depleted isotopic values. The clusters with high HCO_3^- have relatively enriched values for $\delta^{18}O$ and are likely to be older than the clusters with low concentration of HCO_3^- which are similar to young Rhine water. As stated for the FA analysis the high HCO_3^- can be found in the south and western part of the study area. These areas are less influenced by artificial infiltration. Here, the water is very likely older than the young infiltrated Rhine water in the vicinity of the water extraction well gallery.

Figure 6:a) $\delta^{18}O$ - $\delta^{2}H$ [‰] diagram for the study area with the acquired data during the November 2014 sampling campaign. Note that for this campaign only 37 samples were taken and were used for the isotope analysis. In these 37 sampling locations cluster 1 was not present. The black line shows the Global Meteoric Water Line (GMWL) and the dark blue the Local

Meteoric Water Line (LMWL). A different colour indicates the cluster groups. The Birs and Rhine surface water ratio is displayed separately with yellow and blue colours, respectively. The range of Rhine water isotopes is between -11.4 to -10.3 ‰ for $\delta^{18}O$ and -81.8 to -74.1 ‰ for $\delta^{2}H$, respectively. The range of variations for Birs water is -10.1 to -10.14 ‰ for $\delta^{18}O$ and -70.8 to -70.52 ‰ for $\delta^{2}H$, respectively b) $\delta^{18}O$ [‰] - HCO_3^- [mg/l] plot with the six different cluster groups and Birs and Rhine river waters.

4.5 Validation of the multicomponent statistical analysis with organic micropollutants

To further constrain the ongoing processes, persistent organic micropollutants were analysed as a tracer to validate the findings from the aforementioned analysis. We identify the distribution of artificial infiltration compared to the regional groundwater flow in more detail. The use of indicator substances allows to estimate specific pathways (Schirmer et al., 2011; Van Stempvoort et al., 2011). As an example we choose two compounds, namely benzotriazole and sucralose to validate the aforementioned findings and to identify the distribution of artificial infiltration compared to the regional groundwater flow. However, the following described trend for benzotriazole and sucralose is identical for all analysed trace compounds (see also Table S2, supporting information).

According to Giger et al. (2006), benzotriazole is commonly used as a corrosion inhibitor (Sease, 1978), that is fairly water soluble with a limited sorption tendency and shows either no or only slight degradability in ground- and surface water. High benzotriazole concentrations can be found typically in effluent dominated river water, and thus can be used as a tracer for waste water impacted surface water (Moschet et al., 2014; Scheurer et al., 2011). The second compound, sucralose, has been shown to be widespread in the aquatic environment, including waste-, ground- and surface water (Mawhinney et al., 2011). Similar to benzotriazole, sucralose concentrations are typically higher in surface waters than in groundwater.

Both compounds showed a well-defined distribution of their concentration within the clusters (Figure 7). Sampling locations of Cluster C2 close to the infiltration channels and ponds show quite similar concentrations to the concentrations in the river Rhine. The variations in C2 are strongly depending on the distance between the infiltration and sampling location. Sampling locations further away from the infiltration system show smaller concentrations whereas sampling locations directly at the infiltration channels show the highest concentrations. Furthermore, a decreasing trend can be observed also from south-east to north-west along the infiltration channels indicating the channel and pond locations with highest infiltration rates (Figure S13, supporting information). Subsequently, the concentration decreases with increasing cluster number indicating a lesser amount of artificial infiltrated Rhine water. The concentrations of Cluster 6 tend towards the detection limit (benzotriazole 6 ng/l and sucralose 9 ng/l, respectively) for the specific organic micropollutants, indicating that no mixing with surface water occurs.

Figure 7: Trace compound concentration patterns for a) benzotriazole and b) sucralose for Rhine water (input concentration; dark blue square) and the different clusters. Data were acquired during the March 2015 sampling campaign. Note that for this campaign only 43 samples were taken and were used for the analysis. In these 43 sampling locations cluster 1 was not present.

4.6 *Link between clusters and PCE*

In this section, the contamination pattern is related to obtained information about the hydrochemistry for PCE, used as a proxy for most other analysed organic micropollutants. There is a link between the defined cluster numbers associated with different water types and origin and concentration levels for PCE (Figure 8). Here, only small concentrations within a small variability can be found for clusters C1-C3 whereas higher concentrations and a wider variability are displayed for C6. Generally, the mean concentrations increase with increasing

cluster number whereas the increase in cluster number is associated also with higher Ca^{2+} and HCO^{3-} concentrations (see also Figure 5 and 6b).

Figure 8: Boxplot of measured mean concentrations for PCE [µg/l] for all estimated clusters based on the HCA and zoom in for the extraction wells (C1, C2 and C3). N indicates the number of groundwater sampling locations. The coloured filled box shows the lower and upper quartile and the horizontal black line is the median. The whiskers indicate the min and max value without outliers, whereas the circles shows the outliers which are defined as 1.5 times the interquartile range above the upper quartile or below the lower quartile. To validate the robustness of the aforementioned link between the water type and origin with PCE p-values of the t-test between the clusters were calculated to examine whether two clusters are different or not (Table 2). The p-values between C1 and C2 indicate that these clusters are quite similar. The probability is 42%, whereas the probability between C1 and C2 to C3 decreases to 9% and 15%, respectively. The probability that C1 is similar to C5 and C6 within the data set decreases clearly to 2% and 3%, respectively.

Table 2: Calculated p-values of the t-test for different clusters estimated with HCA. A p-value of 1 between clusters indicates that the data sets are 100% equal, whereas a p-value of 0% shows that the data sets are completely different.

p-value	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	1.00					
Cluster 2	0.42	1.00				
Cluster 3	0.09	0.15	1.00			
Cluster 4	-	-	-	1.00	-	-
Cluster 5	0.02	0.03	0.11	-	1.00	
Cluster 6	0.03	0.09	0.29	-	0.36	1.00

Apart from the general trend, the concentrations at the extraction wells show a quite similar trend for PCE. Three different clusters exist (C1, C2 and C3) for the extraction wells. These wells of C3 cluster show slightly higher concentrations for all investigated compounds (Figure 8, smaller panel). However, the compounds can be measured in all wells, although for C1 and C2 the concentrations were lower than for C3.

6. Summary and Conclusions

In this study, a combined approach of multivariate statistical analysis, namely factor analysis (FA) and hierarchical cluster analysis (HCA), interpretation of geochemical process, isotope data and organic micropollutants was used to develop a sound conceptual process understanding for flow and transport. This conceptual process understanding for flow and transport can be used as a tool to ensure the utmost security for drinking water. For the specific study site the following conclusions can be drawn.

Multivariate statistical analysis helps to identify the spatial distribution of different water types. Inter-aquifer mixing in the subsurface is spatially identified, where a higher amount of artificial infiltrated surface water is mixed with water originating from the regional flow pathway in the vicinity of the extraction wells. The mixing signal can be found also in a few specific extraction wells in the western part of the study area which might be linked to fault induced aquifer connectivity. In addition the analysis indicates geochemical dissolution processes in the vicinity of the pumping well gallery where some sampling points show a small difference in the isotope ratios compared to Rhine water. The obtained results of the multivariate statistical analysis are strongly supported by the gained information from isotopic data and organic micropollutant analyses.

Generally, highest concentrations of most contaminants can be found in the south and west whereas a decreasing trend is observed north where the influence of the artificial infiltration

197 increases. Small concentrations of contaminants such as PCE can be found for all extraction
198 wells but slightly higher for cluster group C3. The highest potential risk is given for drinking
199 water supply in the western part of the study area. Here, particular wells with a higher
200 proportion of regional water coming most likely from the south should be further investigated
201 and might be operated differently. Changing abstraction rates or adjusted pumping depths might
202 decrease the contribution of regional water compared to artificial infiltrated Rhine water.

203 In this study, it was demonstrated that systematically combining a comprehensive data set can
204 help to determine the origin of extracted drinking water. Overall, our integrated approach using
205 different kinds of observations in combination with multicomponent statistical analysis, isotope
206 and micropollutant data can be easily transferred to a variety of hydrological settings to
207 synthesise and evaluate large hydrochemical datasets. It is conceivable that the combination
208 with additional data with different information content enables effective interpretation of
209 hydrological processes. Using the applied integrated approach leads to more sound conceptual
210 models which are acting as the very basis to develop improved water resources management
211 practices in a sustainable way.

Acknowledgments

The authors acknowledge the financial support from the Canton Basel-Landschaft, Switzerland in the framework of the project “Regionale Wasserversorgung Basel-Landschaft 21” as well as internal Eawag Discretionary Funding. This study was further supported by the Competence Center Environment and Sustainability (CCES) of the ETH domain in the framework of the RECORD Catchment project (Coupled Ecological, Hydrological and Social Dynamics in Restored and Channelized Corridors of a River at the Catchment Scale). We are grateful for the valuable feedback by the reviewer Nelly Montcoudiol and an additional anonymous reviewer.

References

Aegerter, I., Bosshard, A., 1999. Technischer Bericht-Setzungsproblematik/Massnahmen gegen die Setzung.

Affolter, A., Zechner, E., Huggenberger, P., 2010. Grundwassermodell Unteres Birstal - Rhein - Muttenz Evaluation der Zuströmbereiche der Trinkwasserfassungen Muttenz und Hardwasser AG Technischer Bericht, BGA BL-1.

Atchley, A.L., Navarre-Sitchler, A.K., Maxwell, R.M., 2014. The effects of physical and geochemical heterogeneities on hydro-geochemical transport and effective reaction rates. J Contam Hydrol, 165: 53-64.

Auckenthaler, A., Baenninger, D., A., A., Zechner, E., Huggenberger, P., 2010. Drinking water production close to contaminant sites: a case study from the region of Basel, Switzerland. GQ10: Groundwater Quality Management in a Rapidly Changing World (Proc. 7th International Groundwater Quality Conference held in Zurich, Switzerland, 13–18 June 2010). IAHS Publ 342, 2011, 167-170.

Avon, L., Bredehoeft, J.D., 1989. An Analysis of Trichloroethylene Movement in Groundwater at Castle-Air-Force-Base, California. J Hydrol, 110(1-2): 23-50.

236 Bosma, T.N.P. et al., 1994. Comparison of Reductive Dechlorination of Hexachloro-1,3-Butadiene in
237 Rhine Sediment and Model Systems with Hydroxocobalamin. Environ Sci Technol, 28(6): 1124-
238 1128.
239 Chen, L.W., Gui, H.R., Yin, X.X., 2011. Monitoring of flow field based on stable isotope geochemical
240 characteristics in deep groundwater. Environ Monit Assess, 179(1-4): 487-498.
241 Cloutier, V., Lefebvre, R., Therrien, R., Savard, M.M., 2008. Multivariate statistical analysis of
242 geochemical data as indicative of the hydrogeochemical evolution of groundwater in a
243 sedimentary rock aquifer system. J Hydrol, 353(3-4): 294-313.
244 Daughney, C.J., Raiber, M., Moreau-Fournier, M., Morgenstern, U., van der Raaij, R., 2012. Use of
245 hierarchical cluster analysis to assess the representativeness of a baseline groundwater quality
246 monitoring network: comparison of New Zealand's national and regional groundwater
247 monitoring programs. Hydrogeol J, 20(1): 185-200.
248 Davis, J.C., 1986. Statistic and Data Analysis in Geology. John Wiley and Sons INC., New York.
249 Demlie, M., Wohnlich, S., Wisotzky, F., Gizaw, B., 2007. Groundwater recharge, flow and
250 hydrogeochemical evolution in a complex volcanic aquifer system, central Ethiopia. Hydrogeol
251 J, 15(6): 1169-1181.
252 Dilsiz, C., 2006. Conceptual hydrodynamic model of the Pamukkale hydrothermal field, southwestern
253 Turkey, based on hydrochemical and isotopic data. Hydrogeol J, 14(4): 562-572.
254 Diamond, M.L., Hodge, E., 2007. Urban Contaminant Dynamics: From Source to Effect. Environ Sci
255 Technol, 41(11): 3796-3800. DOI:10.1021/es072542n
256 Doherty, J., Simmons, C.T., 2013. Groundwater modelling in decision support: reflections on a unified
257 conceptual framework. Hydrogeol J, 21(7): 1531-1537.
258 Doherty, R.E., 2000. A history of the production and use of carbon tetrachloride, tetrachloroethylene,
259 trichloroethylene and 1,1,1-trichloroethane in the United States: Part 1 - Historical
260 background; Carbon tetrachloride and tetrachloroethylene. Environ Forensics, 1(2): 69-81.

Fields, J.A.u.S.-A., R, 2004. Biodegradability of chlorinated solvents and related chlorinated aliphatic compounds. Science dossier, Euro Chlor, University of Arizonano Letters.

Gabriel, T., Meier, T., 2014. Der Aktivkohlefilter Hard-Die neue Anlage der Hardwasser AG. AQUA & GAS, Nr. 12.

Gandhi, R.K., Hopkins, G.D., Goltz, M.N., Gorelick, S.M., McC rty, P.L., 2002a. Full-scale demonstration of in situ cometabolic biodegradation of trichloroethylene in groundwater - 1. Dynamics of a recirculating well system. Water Resour Res, 38(4).

Gandhi, R.K., Hopkins, G.D., Goltz, M.N., Gorelick, S.M., McCarty, P.L., 2002b. Full-scale demonstration of in situ cometabolic biodegradation of trichloroethylene in groundwater - 2. Comprehensive analysis of field data using reactive transport modeling. Water Resour Res, 38(4).

Giger, W., Schaffner, C., Kohler, H.P.E., 2006. Benzotriazole and tolyltriazole as aquatic contaminants. 1. Input and occurrence in rivers and lakes. Environ Sci Technol, 40(23): 7186-7192.

Güler, C., Thyne, G., McCray, J., Turner, K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol J, 10(4): 455-474. DOI:10.1007/s10040-002-0196-6

Gürler, B., Hauber, L., Schwander, M., 1987. Die Geologie der Umgebung von Basel mit Hinweisen übder die Nutzungsmöglichkeiten der Erdwärme. Beitrag zur Geologischen Karte der Schweiz.

Hothorn, T., Everitt, B., 2009. A Handbook of Statistical Analyses Using R, Second Edition. Chapman and Hall.

Huggenberger, P., Affolter, A., Zechner, E., Dresma, H., 2009. Stationär kalibriertes Grundwassermodell Muttenz unteres Birstal, Berechnung geschichtlicher Szenarien zur Abschätzung der Schadstoffverteilung der Deponien in Muttenz. Geologisches Institut der Universität Basel.

Hunt, R.J., Doherty, J., Tonkin, M.J., 2007. Are models too simple? Arguments for increased parameterization. Ground Water, 45(3): 254-262.

285 Huntscha, S., Velosa, D.M.R., Schroth, M.H., Hollender, J., 2013. Degradation of Polar Organic
286 Micropollutants during Riverbank Filtration: Complementary Results from Spatiotemporal
287 Sampling and Push-Pull Tests. Environ Sci Technol, 47(20): 11512-11521.
288 Hynds, P., Misstear, B.D., Gill, L.W., Murphy, H.M., 2014. Groundwater source contamination
289 mechanisms: Physicochemical profile clustering, risk factor analysis and multivariate
290 modelling. J Contam Hydrol, 159: 47-56.
291 King, A.C., Raiber, M., Cox, M.E., 2014. Multivariate statistical analysis of hydrochemical data to assess
292 alluvial aquifer-stream connectivity during drought and flood: Cressbrook Creek, southeast
293 Queensland, Australia. Hydrogeol J, 22(2): 481-500.
294 Koh, D.C. et al., 2009. Baseline geochemical characteristics of groundwater in the mountainous area of
295 Jeju Island, South Korea: Implications for degree of mineralization and nitrate contamination.
296 J Hydrol, 376(1-2): 81-93.
297 Koltermann, C.E., Gorelick, S.M., 1996. Heterogeneity in sedimentary deposits: A review of structure-
298 imitating, process-imitating, and descriptive approaches. Water Resour Res, 32(9): 2617-2658.
299 Levison, J., Novakowski, K., Reiner, E., Kolic, T., 2012. Potential of groundwater contamination by
300 polybrominated diphenyl ethers (PBDEs) in a sensitive bedrock aquifer (Canada). Hydrogeol J,
301 20(2): 401-412. DOI:10.1007/s10040-011-0813-3
302 Levison, J.K., Novakowski, K.S., 2012. Rapid transport from the surface to wells in fractured rock: A
303 unique infiltration tracer experiment. J Contam Hydrol, 131(1–4): 29-38.
304 DOI:http://dx.doi.org/10.1016/j.jconhyd.2012.01.001
305 Loague, K., Abrams, R.H., Davis, S.N., Nguyen, A., Stewart, I.T., 1998. A case study simulation of DBCP
306 groundwater contamination in Fresno County, California - 2. Transport in the saturated
307 subsurface. J Contam Hydrol, 29(2): 137-163.
308 Mackay, D.M., Roberts, P.V., Cherry, J.A., 1985. Transport of Organic Contaminants in Groundwater.
309 Environ Sci Technol, 19(5): 384-392.

Mawhinney, D.B., Young, R.B., Vanderford, B.J., Borch, T., Snyder, S.A., 2011. Artificial Sweetener Sucralose in U.S. Drinking Water Systems. Environ Sci Technol, 45(20): 8716-8722.

Mayer, K.U., Blowes, D.W., Frind, E.O., 2001. Reactive transport modeling of an in situ reactive barrier for the treatment of hexavalent chromium and trichloroethylene in groundwater. Water Resour Res, 37(12): 3091-3103.

MBN, 2008. Hydrogeologische Verhältnisse im Gebiet der Trinkwassergewinnung Hardwald (Muttenz, Kt. BL). Bericht W1482B, https://www.baselland.ch/fileadmin/baselland/files/docs/bud/aue/grundwasser/form/mbn_hardwald_2008_bericht.pdf.

McCallum, J.L., Cook, P.G., Simmons, C.T., 2015. Limitations of the Use of Environmental Tracers to Infer Groundwater Age. Groundwater, 53: 56-70.

Messier, K.P., Akita, Y., Serre, M.L., 2012. Integrating Address Geocoding, Land Use Regression, and Spatiotemporal Geostatistical Estimation for Groundwater Tetrachloroethylene. Environ Sci Technol, 46(5): 2772-2780.

Moeck, C., Hunkeler, D., Brunner, P., 2015. Tutorials as a flexible alternative to GUIs: An example for advanced model calibration using Pilot Points. Environ Modell Softw, 66: 78-86.

Montcoudiol, N., Molson, J., Lemieux, J.M., 2015. Groundwater geochemistry of the Outaouais Region (Quebec, Canada): a regional-scale study. Hydrogeol J, 23(2): 377-396.

Morgenstern, U. et al., 2015. Using groundwater age and hydrochemistry to understand sources and dynamics of nutrient contamination through the catchment into Lake Rotorua, New Zealand. Hydrol Earth Syst Sc, 19(2): 803-822.

Moschet, C. et al., 2014. How a Complete Pesticide Screening Changes the Assessment of Surface Water Quality. Environ Sci Technol, 48(10): 5423-5432.

Moya, C.E., Raiber, M., Taulis, M., Cox, M.E., 2015. Hydrochemical evolution and groundwater flow processes in the Galilee and Eromanga basins, Great Artesian Basin, Australia: A multivariate statistical approach. Sci Total Environ, 508: 411-426.

Murphy, S., Ouellon, T., Ballard, J.M., Lefebvre, R., Clark, I.D., 2011. Tritium-helium groundwater age used to constrain a groundwater flow model of a valley-fill aquifer contaminated with trichloroethylene (Quebec, Canada). Hydrogeol J, 19(1): 195-207.

Nagra, 2002. Project Opalinuston: syntheses of geoscientific investigation results (in German). NTB (2002): pp. 02–03.

Ninomiya, K., Sakai, M., Ohba, E., Kashiwagi, N., 1994. Kinetic-Model for the Biotransformation of Tetrachloroethylene in Groundwater. Water Sci Technol, 30(7): 13-18.

Obeidat, M.M., Awawdeh, M., Abu Al-Rub, F., 2013. Multivariate statistical analysis and environmental isotopes of Amman/Wadi Sir (B2/A7) groundwater, Yarmouk River Basin, Jordan. Hydrol Process, 27(17): 2449-2461.

Panda, U.C., Sundaray, S.K., Rath, P., Nayak, B.B., Bhatta, D., 2006. Application of factor and cluster analysis for characterization of river and estuarine water systems - A case study: Mahanadi River (India). J Hydrol, 331(3-4): 434-445.

Parkhurst, D., Appelo, C., 1999. User's guide to PHREEQC (version 2): a computer program for speciation, batch-reaction,one-dimensional transport, and inverse geochemical calcula-tions. US Geol Surv Water Resour Invest Rep 99–4259, 326 pp.

Poulsen, M.M., Kueper, B.H., 1992. A Field Experiment to Study the Behavior of Tetrachloroethylene in Unsaturated Porous-Media. Environ Sci Technol, 26(5): 889-895.

Scheurer, M. et al., 2011. Correlation of six anthropogenic markers in wastewater, surface water, bank filtrate, and soil aquifer treatment. J Environ Monitor, 13(4): 966-973.

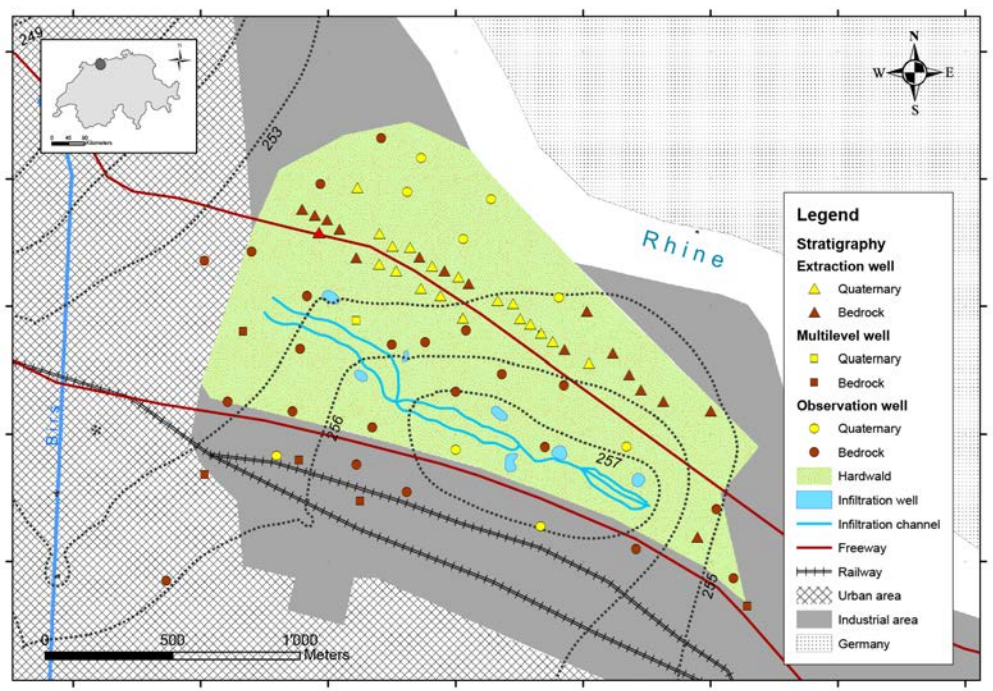
Schirmer, M., Leschik, S., Musolff, A., 2013. Current research in urban hydrogeology - A review. Adv Water Resour, 51: 280-291.

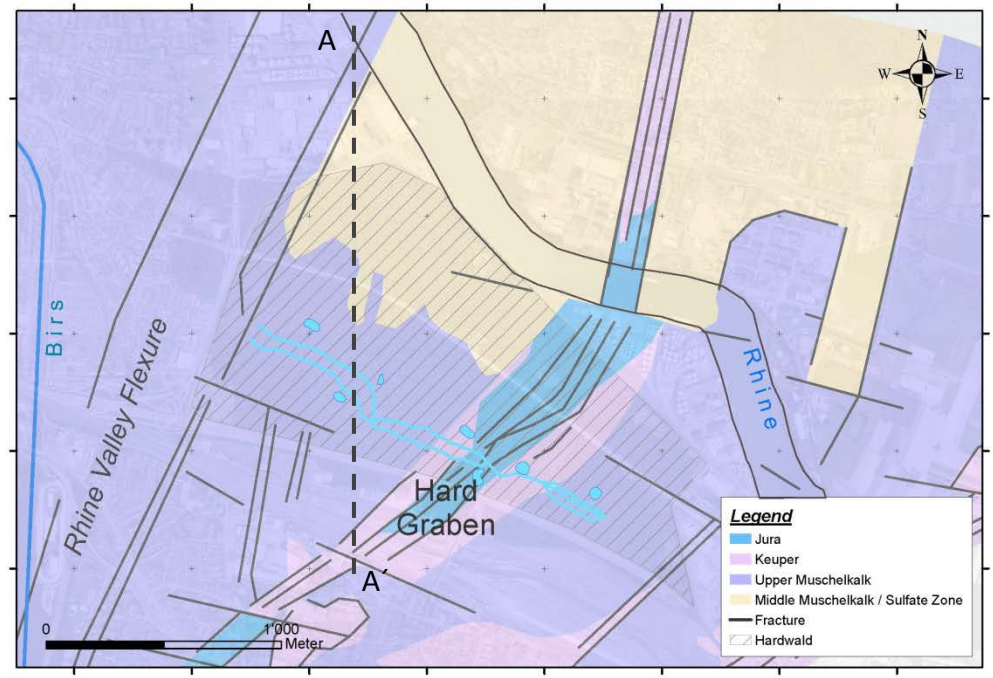
Schirmer, M. et al., 2011. Mass fluxes of xenobiotics below cities: challenges in urban hydrogeology. Environ Earth Sci, 64(3): 607-617.

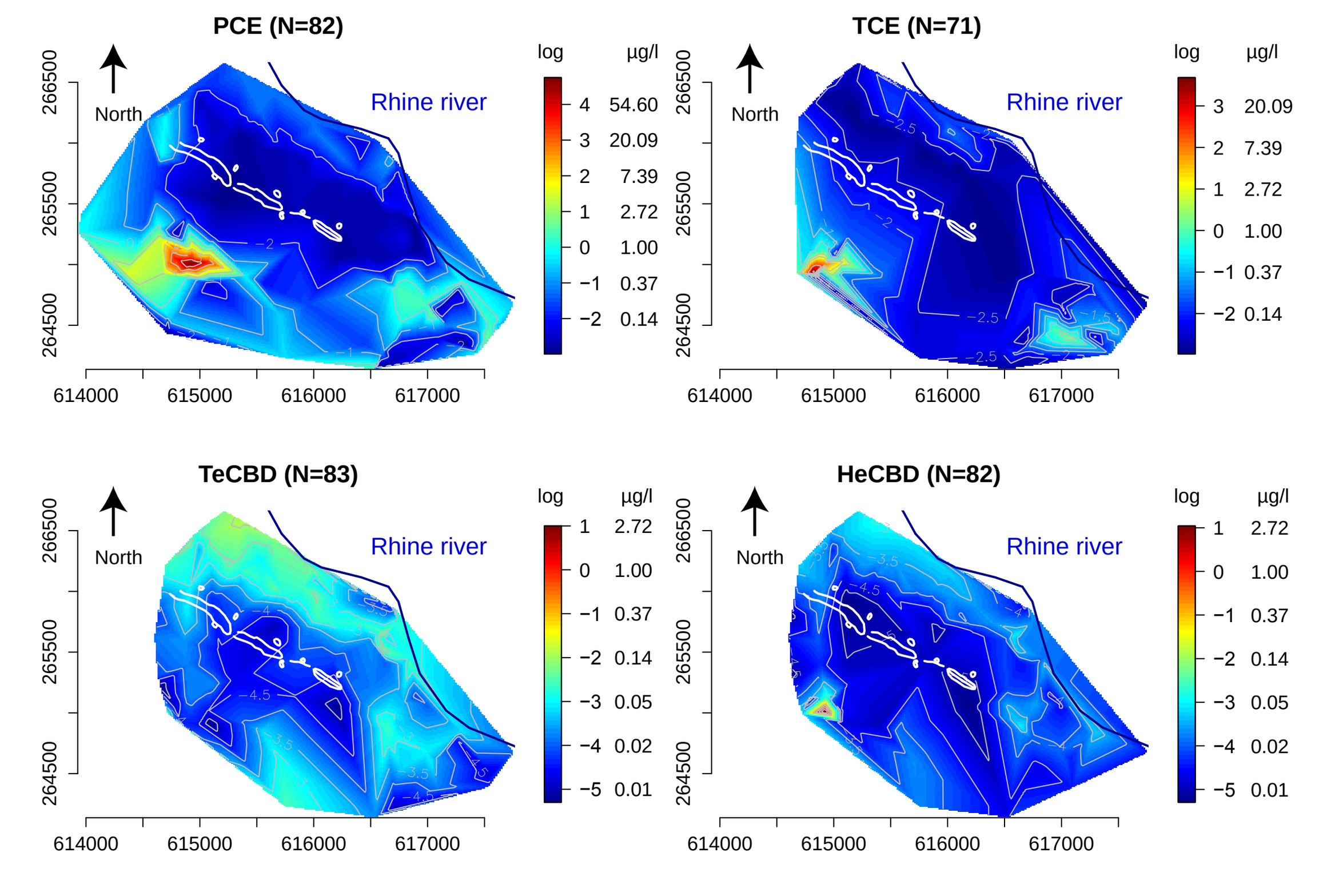
Sease, C., 1978. Benzotriazole: A Review for Conservators. Studies in Conservation, 23(2): 76-85. DOI:10.2307/1505798

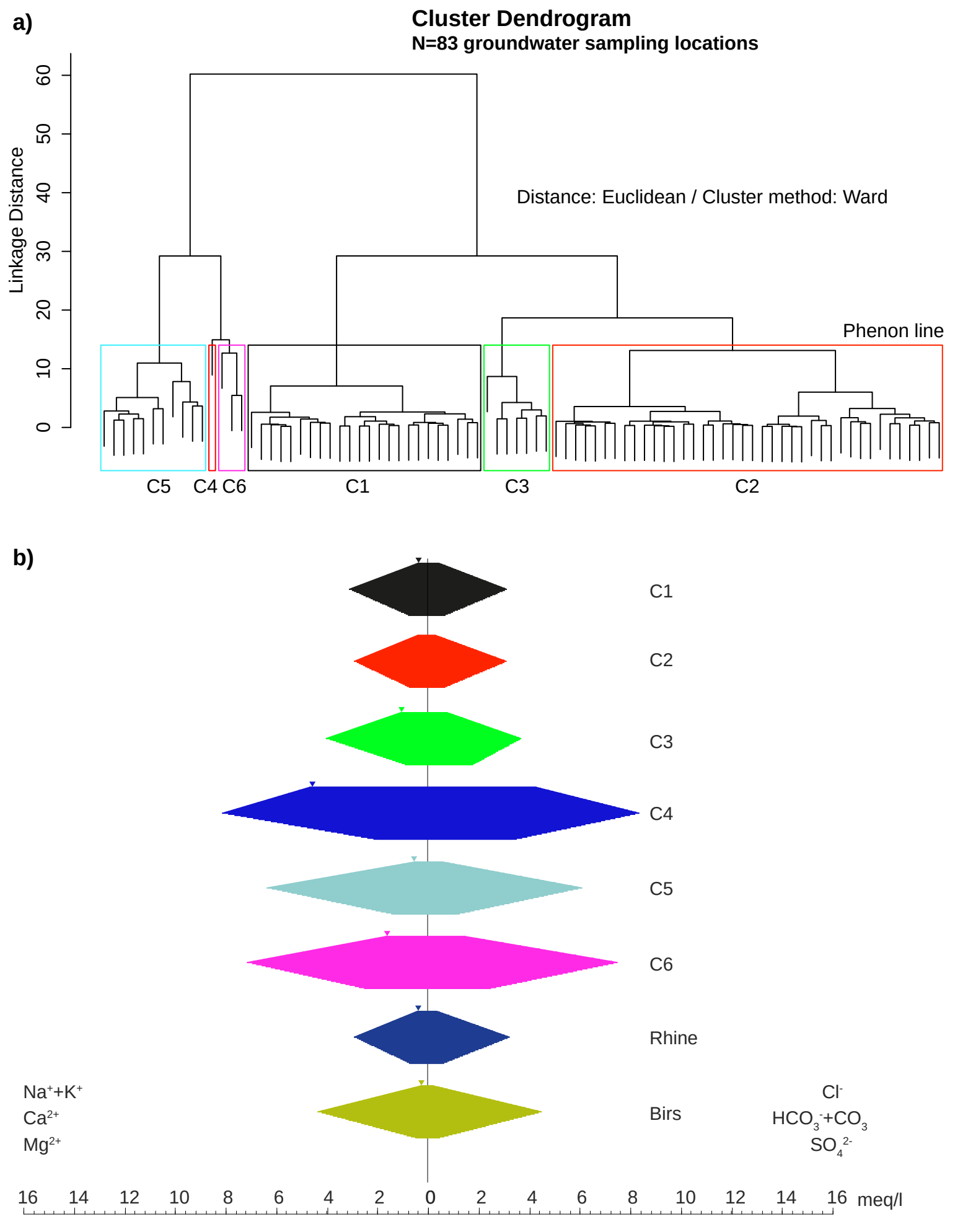
362 Selle, B., Schwientek, M., Lischeid, G., 2013. Understanding processes governing water quality in
363 catchments using principal component scores. J Hydrol, 486: 31-38.
364 Serrano, S.E., 2003. Propagation of nonlinear reactive contaminants in porous media. Water Resour
365 Res, 39(8).
366 Singh, U.K. et al., 2008. Assessment of the impact of landfill on groundwater quality: A case study of
367 the Pirana site in western India. Environ Monit Assess, 141(1-3): 309-321.
368 Spottke, I., Zechner, E., Huggenberger, P., 2005. The southeastern border of the Upper Rhine Graben:
369 a 3D geological model and its importance for tectonics and groundwater flow. Int J Earth Sci,
370 94(4): 580-593.
371 Stieglitz, L., Roth, W., Kühn, W., Leger, W., 1976. Das Verhalten von Organohalogenverbindungen. Vom
372 Wasser, 47: 347-377.
373 Suk, H., Lee, K.-K., 1999. Characterization of a Ground Water Hydrochemical System Through
374 Multivariate Analysis: Clustering into Ground Water Zones. Ground Water, 37(3): 358-366.
375 DOI:10.1111/j.1745-6584.1999.tb01112.x
376 Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical
377 clustering. Bioinformatics, 22(12): 1540-1542.

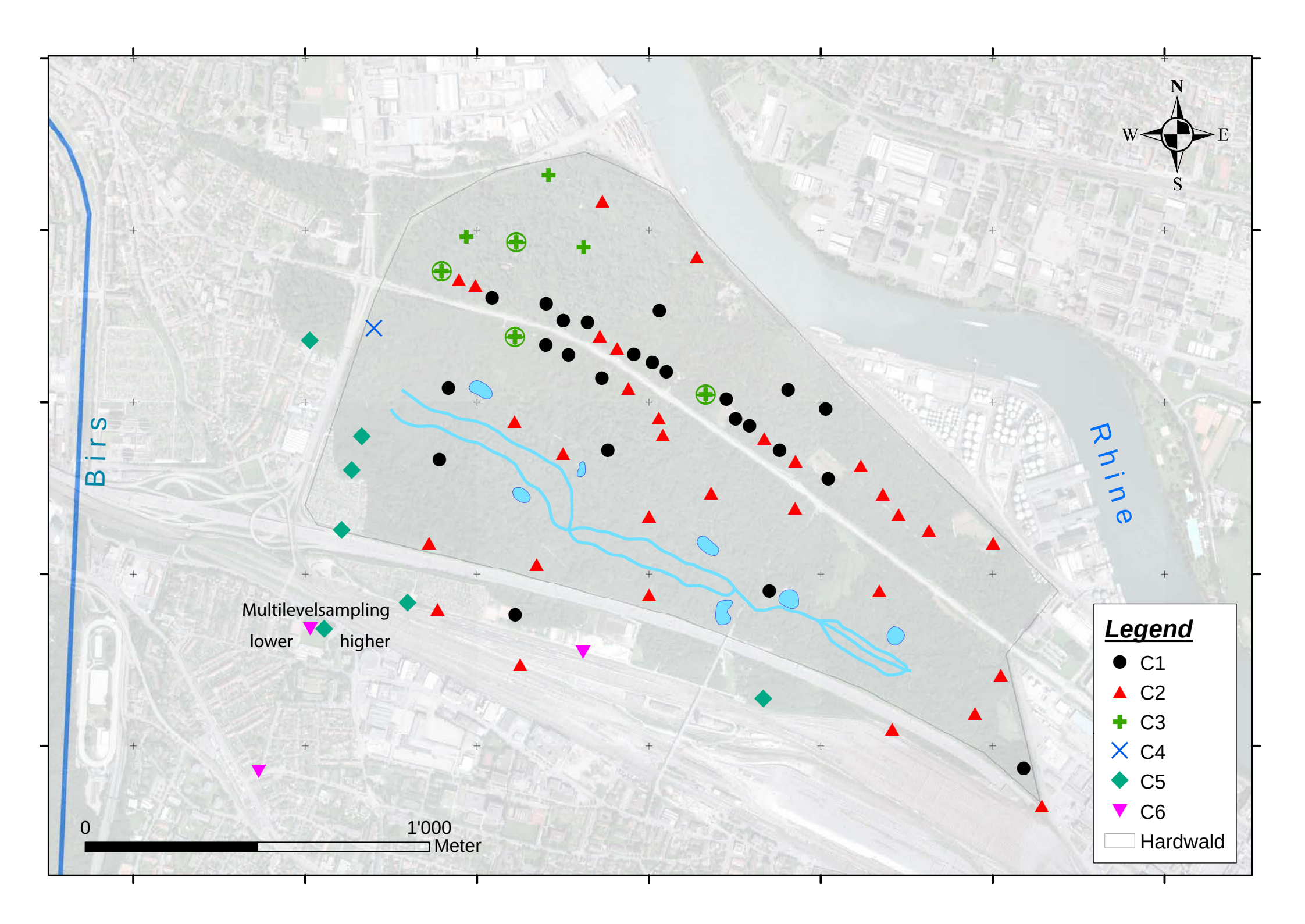
378 Thyne, G., Guler, C., Poeter, E., 2004. Sequential analysis of hydrochemical data for watershed
379 characterization. Ground Water, 42(5): 711-723.
380 van den Brink, C., Frapporti, G., Griffioen, J., Zaadnoordijk, W.J., 2007. Statistical analysis of
381 anthropogenic versus geochemical-controlled differences in groundwater composition in The
382 Netherlands. J Hydrol, 336(3-4): 470-480.
383 Van Stempvoort, D.R., Roy, J.W., Brown, S.J., Bickerton, G., 2011. Artificial sweeteners as potential
384 tracers in groundwater in urban environments. J Hydrol, 401(1–2): 126-133.
385 DOI:http://dx.doi.org/10.1016/j.jhydrol.2011.02.013


249
253
256
257
245
Rhine
Birs
N
W
E
S
Legend
Stratigraphy
Extraction well
Quaternary
Bedrock
Multilevel well
Quaternary
Bedrock
Observation well
Quaternary
Bedrock
Hardwald
Infiltration well
Infiltration channel
Freeway
Railway
Urban area
Industrial area
Germany
0
500
1'000
Meters


A
A´
N
W
E
S
Birs
Rhine Valley Flexure
Rhine
Hard
Graben
Legend
Jura
Keuper
Upper Muschelkalk
Middle Muschelkalk / Sulfate Zone
Fracture
Hardwald
0
1'000
Meter


PCE (N=82)
North
Rhine river
log
µg/l
4 54.60
3 20.09
2 7.39
1 2.72
0 1.00
−1 0.37
−2 0.14
266500
265500
264500
614000
615000
616000
617000
TCE (N=71)
North
Rhine river
log
µg/l
3 20.09
2 7.39
1 2.72
0 1.00
−1 0.37
−2 0.14
266500
265500
264500
614000
615000
616000
617000
TeCBD (N=83)
North
Rhine river
log
µg/l
1 2.72
0 1.00
−1 0.37
−2 0.14
−3 0.05
−4 0.02
−5 0.01
266500
265500
264500
614000
615000
616000
617000
HeCBD (N=82)
North
Rhine river
log
µg/l
1 2.72
0 1.00
−1 0.37
−2 0.14
−3 0.05
−4 0.02
−5 0.01
266500
265500
264500
614000
615000
616000
617000


a)
Cluster Dendrogram
N=83 groundwater sampling locations
Linkage Distance
0
10
20
30
40
50
60
Distance: Euclidean / Cluster method: Ward
Phenon line
C5
C4
C6
C1
C3
C2
b)
C1
C2
C3
C4
C5
C6
Rhine
Birs
Na+K
Ca2
Mg2
Cl
HCO3+CO3
SO4 2-
16 14 12 10 8 6 4 2 0 2 4 6 8 10 12 14 16
meq/l


N
W
E
S
Birs
Rhine
Multilevelsampling
lower
higher
0
1'000
Meter
Legend
C1
C2
C3
C4
C5
C6
Hardwald

Legend
C1
C2
C3
C4
C5
C6

Theoretical line Calcite dissolution

$[Ca^{2+}]=0.5[HCO_3^-]$

Birs river water

Rhine river water

Ca^{2+} (mmol/l)

HCO_3^- (mmol/l)

0
1
2
3
4
5
6

0
2
4
6
8

a)

Legend
C2
C3
C4
C5
C6

GMWL
LMWL
Birs river water
Rhine river water

^{2}H (‰)
−65
−70
−75
−80
−12
−11
−10
−9
−8

^{18}O (‰)

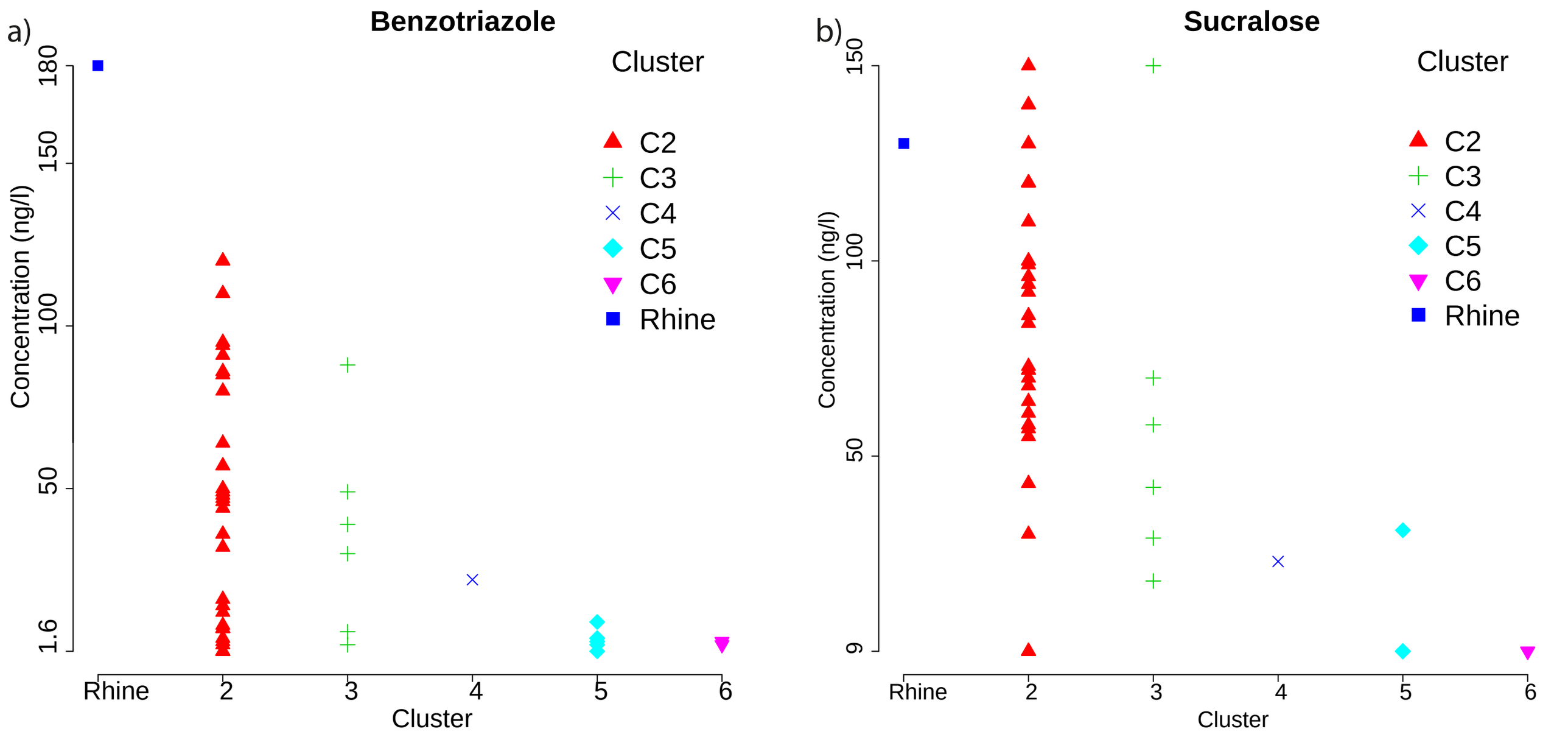
b)

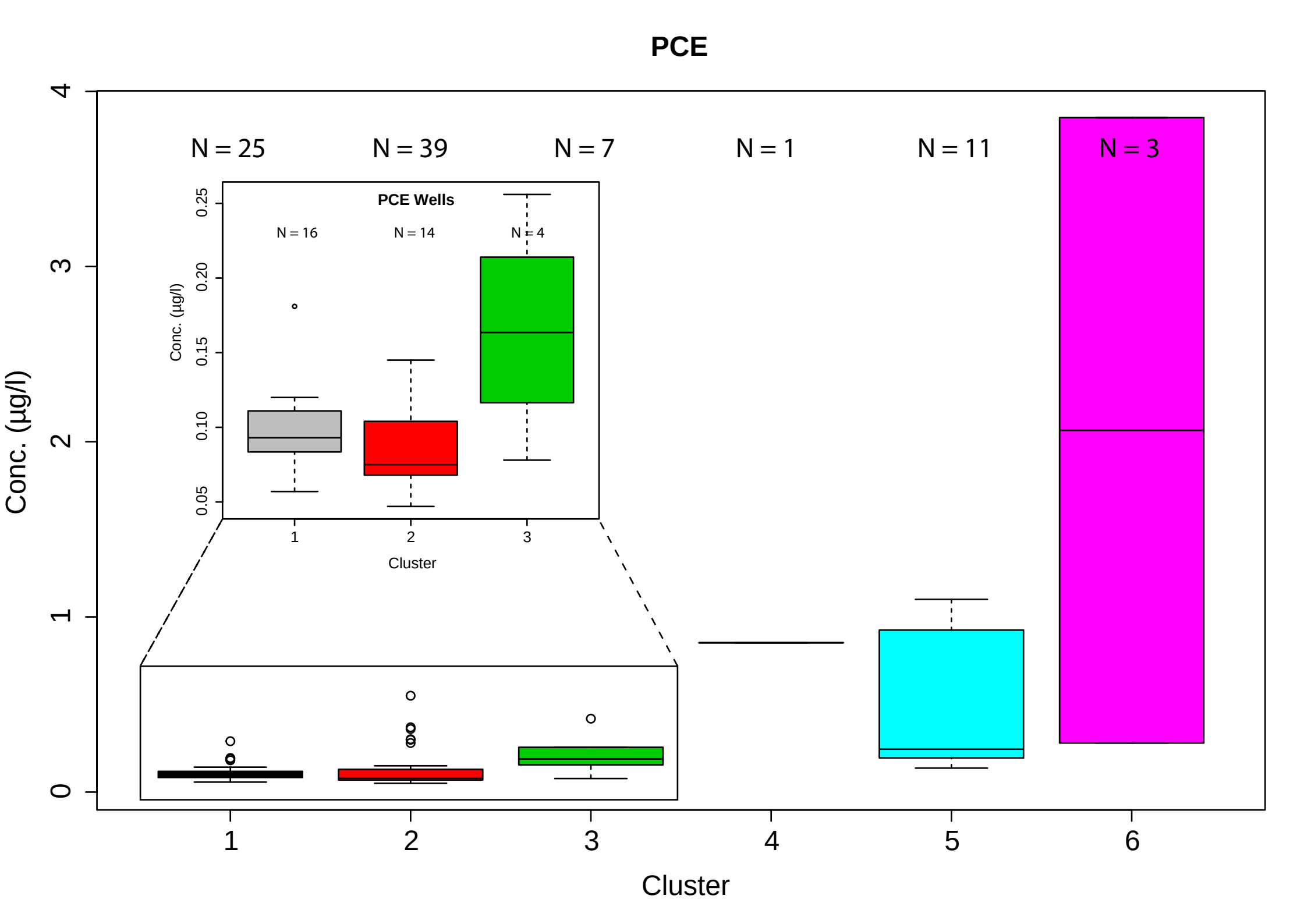
Legend
C2
C3
C4
C5
C6

Birs river water
Rhine water

HCO_3^- (mg/l)
500
450
400
350
300
250
200
−11.0
−10.5
−10.0
−9.5
−9.0

^{18}O (‰)


a)
Benzotriazole
Concentration (ng/l)
180
150
100
50
1.6
Cluster
C2
C3
C4
C5
C6
Rhine
Rhine
2
3
4
5
6
Cluster
b)
Sucralose
Concentration (ng/l)
150
100
50
9
Cluster
C2
C3
C4
C5
C6
Rhine
Rhine
2
3
4
5
6
Cluster


PCE
N = 25
N = 39
N = 7
N = 1
N = 11
N = 3
Conc. (µg/l)
0
1
2
3
4
1
2
3
4
5
6
Cluster
PCE Wells
N = 16
N = 14
N = 4
Conc. (µg/l)
0.05
0.10
0.15
0.20
0.25
1
2
3
Cluster