Research article

Geoenergy

https://doi.org/10.1144/geoenergy2023-036 | Vol. 2 | 2024 | geoenergy2023-036

Fingerprinting dissolved organic compounds: a potential tool for identifying the surface infiltration environments of meteoric groundwaters

M. Stillings[1]*, R. J. Lunn[1], Z. K. Shipton[1], R. A. Lord[1], S. Thompson[2] and M. Knapp[1]

[1] Department of Civil & Environmental Engineering, University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, UK
[2] Nuclear Waste Services, Pelham House, Pelham Drive, Calderbridge, Cumbria CA20 1DB, UK
MS, 0000-0001-8418-6866
* Correspondence: mark.stillings@strath.ac.uk

Abstract: Current methods for tracing decades-old groundwaters rely on isotope geochemistry to determine groundwater age and altitude at the point of infiltration. Temporal and spatial variability in atmospheric conditions, and water–rock interactions, can make the interpretation of isotopes uncertain. Here, we propose a new method of groundwater tracing based on the fingerprinting of natural dissolved organics. We present our initial findings from the Grimsel Test Site in Switzerland, located within a fractured granite. Using 2D gas chromatography, we derive detailed organic fingerprints from surface soils at several locations and show that different soils produce distinctly different dissolved organic signatures. We then compare the soils with groundwater and lake water using a non-targeted approach employing principal component analysis and hierarchical cluster analysis. Our analysis finds three statistically significant clusters. Most groundwaters are clustered with the lake-water samples but two are clustered with soil from the highest altitude surface sampling location. We hypothesize that for samples to form a significant cluster, they must have been derived from a common environment, with a unique combination of organic compounds. For groundwaters to cluster with soil samples or lake water, we theorize there must be a hydraulic connection between the type of infiltration environment and the groundwater sampling locations within each cluster. Our research demonstrates that organic molecules derived from the surface environment can be used to discriminate near-surface environment(s) through which meteoric groundwater has infiltrated. Organic fingerprinting could prove a powerful tool for improved understanding of groundwater flow systems, particularly when combined with other complementary techniques.

Supplementary material: Compound alignment data set and supplementary tables are available at https://doi.org/10.6084/m9.figshare.c.7129987

Thematic collection: This article is part of the Sustainable geological disposal and containment of radioactive waste collection available at: https://www.lyellcollection.org/topic/collections/radioactive

Received 27 September 2023; **revised** 25 January 2024; **accepted** 12 February 2024

The chemical composition of groundwater, due to both natural and man-made processes, has long been used for tracing the origins and ages of subsurface waters. Groundwater tracers generally fall into three types: (1) additive tracers/point-source pollutants (Flury and Wai 2003; Abrantes *et al.* 2018), which disperse rapidly and may be unrecoverable (Filippini *et al.* 2018), and hence can only be used to trace flow over short distances and timescales; (2) tritium, helium-4 and CFCs, which can be related to specific time points (the bomb pulse tests in the 1950s and the banning of CFCs in the 1980s) to determine the age of modern meteoric groundwaters (Casillas-Trasvina *et al.* 2022; Okuhata *et al.* 2022), although tritium concentrations have now decayed to barely detectible levels; and (3) isotope ratios, which can be used to determine the original altitude of groundwater infiltration (e.g. $\delta^{18}O$ and δD isotopes) (Prada *et al.* 2016; Schneeberger *et al.* 2017; Fackrell *et al.* 2020), the presence of differing hydrothermal and lithological water sources (e.g. $\delta^{18}O(SO_4)$, $\delta^{34}S(SO_4)$, $\delta^{37}Cl$, ^{3}H, $\delta^{14}C$ and ^{87}Sr) (Pichler 2005; Osman Awaleh *et al.* 2020), and the mixed origins of old groundwaters based on age (e.g. ^{3}He, ^{39}Ar, ^{81}Kr and ^{85}Kr) (Kralik *et al.* 2014; Gerber *et al.* 2017; Avrahamov *et al.* 2018). Vascular plant biomarkers can be used to determine the origin of dissolved organic matter from near-surface environments (Shen *et al.* 2015). Most hydrogeological studies use a combination of several tracing techniques to reduce uncertainty in the determination of groundwater origins. Using existing techniques, meteoric waters can be identified, their ages predicted and, where topographical elevation varies, the altitude of infiltration can be estimated. However, for most groundwaters the type of near-surface infiltration environment (e.g. surface-soil type, river bed or lake bed) through which the groundwater infiltrated cannot be reliably determined.

Natural dissolved organic compounds are not routinely used for groundwater tracing, although they are prevalent in all aquatic environments, including groundwater systems. Dissolved organic compounds are input into the groundwater system through the breakdown of solid organic matter in the form of organisms (Shen *et al.* 2015) and soil/plant litter (Baker *et al.* 2000), or through the transport of water-soluble organic compounds and pollutants (Khatri and Tyagi 2015) from the surface. Other sources within groundwater include organic matter within the host rock and excretions from subsurface micro-organisms. Through time phyto-, microbial- and chemical- degradation (Chen *et al.* 2010; Obernosterer and Benner 2004; Zhang *et al.* 2009) result in a breakdown of larger solid organic compounds into a series of smaller water-soluble compounds. Breakdown of solid organic matter is rapid at the surface where oxygen is readily available to aid in biotic degradation (Keiluweit *et al.* 2016). Infiltrating meteoric water carries these water-soluble organic compounds into groundwater. Once in the groundwater, decay continues to alter organic compound structure and composition but the rate of decay decreases substantially with increasing depth, due to the increasingly anaerobic conditions

(Kortelainen and Karhu 2006). Hence, groundwaters contain a complex array of preserved dissolved organic compounds.

The tracing of groundwater using dissolved organic carbon has seen previous success using a targeted approach (Derrien *et al.* 2017). Until recent years, the detection, measurement and comparison of the complete organic molecular composition of a water sample, and the relative abundance of individual molecules through a non-targeted approach, has not been readily achievable. However, with the advent of 2D gas chromatography time of flight mass spectrometry (GC × GC-ToF-MS) (Patrushev 2015), detailed organic fingerprinting of water samples is now possible. GC × GC is largely confined to the fields of environmental forensics, where it is used as legal evidence of the relative contributions of individual polluters (McGregor *et al.* 2012; Amaral *et al.* 2020), and medical sciences, where minute changes in the organic composition of bodily fluids could provide an early indication of disease (Almstetter *et al.* 2012).

In this research, we show that differing near-surface infiltration environments (individual soil types and lakes) have distinct dissolved organic signatures and that these signatures can be detected within groundwater samples at depth. We collected samples from a number of surface sites and groundwater samples from multiple boreholes at the Grimsel Test Site, Switzerland. These signatures were then compared visually using principal component analysis (PCA) and placed into statistically significant groups identified through hierarchical cluster analysis (HCA). We show that specific near-surface infiltration environments have differing organic signatures. Further, these organic signatures are distinguishable in groundwater samples at depth and, hence, could be used to indicate the predominant near-surface infiltration environment. Our research demonstrates that organic fingerprinting may prove a useful investigative tool for distinguishing the dominant near-surface infiltration environment(s) through which individual groundwater samples have infiltrated.

Field site

The Grimsel Test Site (GTS) is located in the Hasli Valley in the Canton of Bern (Switzerland), and comprises a series of access tunnels and groundwater monitoring boreholes (marked in black in Fig. 1). The entrance tunnel is at the base of the reservoir dam and the GTS site is *c.* 30 m below the reservoir bed, and between 200 and 500 m below the ground surface (which slopes steeply downwards from south to north). Boreholes and tunnels cut two lithologies: the Central Aar Granite (CAGr) to the north and the Grimsel Granodiorite (GrGr) to the south. The fracture network comprises open (unfilled) and gouge-filled fractures; fracture flow dominates the groundwater system at the GTS (Schneeberger *et al.* 2016).

At the GTS, potential inflows into the groundwater system are from the infiltration of precipitation (rainfall or snow melt). Meteoric waters infiltrate through surface soils, mountain stream beds and from the reservoir beds (where the reservoir level is above the adjacent groundwater head). We hypothesized that different near-surface water infiltration environments (soils and lake water) at the GTS could give rise to different exposure to potential organic solutes. The surface exposure directly above the GTS varies in slope gradient, orientation and altitude, and largely comprises weathered granite with soil and vegetation-filled fissures. Where soil is present, there are clearly visible spatial variations in soil type. On the east-facing mountain side overlying the GTS there are also several

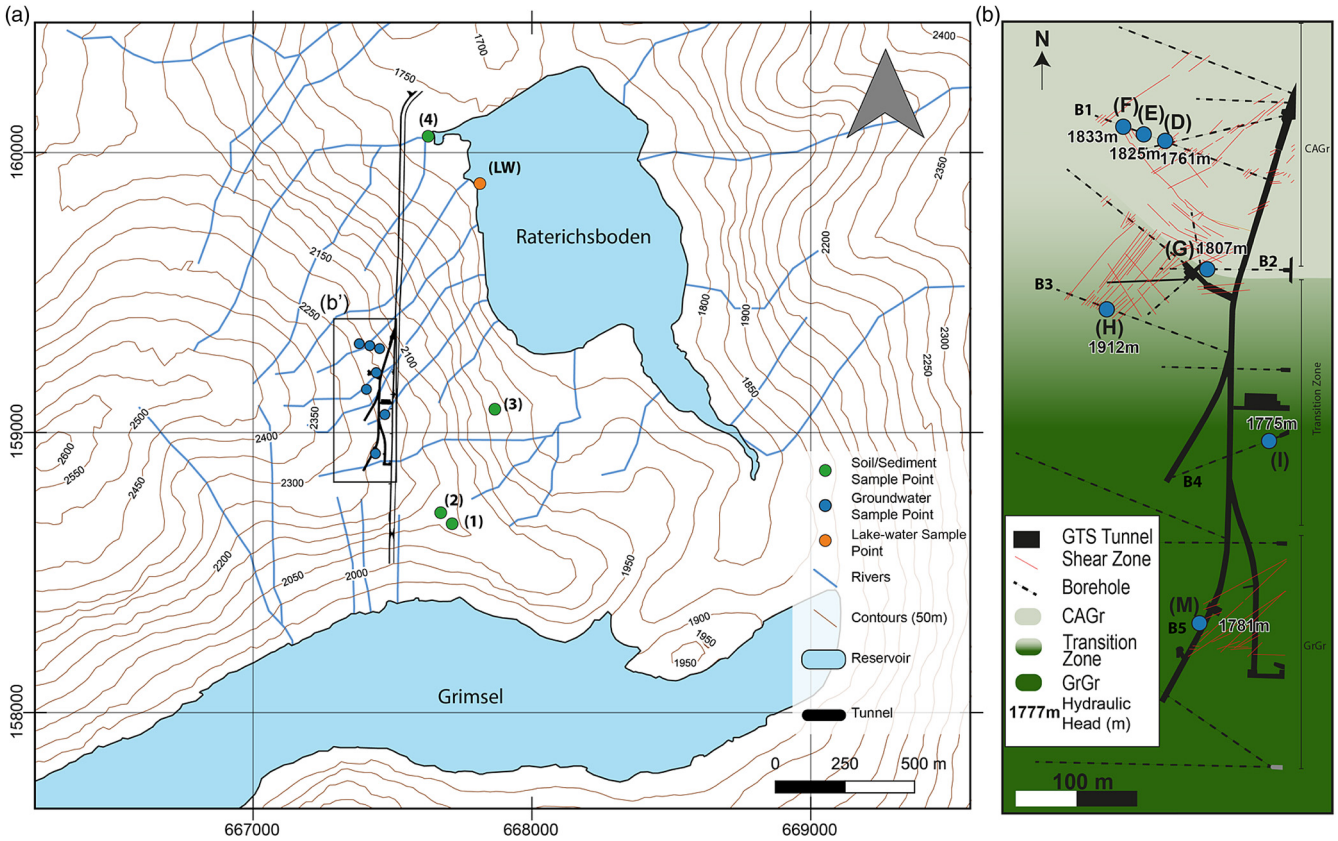

Fig. 1. (**a**) Schematic map of the Grimsel area, including hydroelectric reservoirs (blue), access and GTS tunnels (black), surface soil/sediment sample sites 1–4 (green circles), groundwater sample sites (blue circles) and lake-water sample site (orange circle). (**b**) An expanded schematic map of area (b′) depicted in (a). The map (b′) highlights the groundwater sample locations from each borehole (labelled B1–B5) and the lithological units at 1730 m above mean sea level (m amsl). CAGr, Central Aar Granite (light green); GrGr, Grimsel Granodiorite (dark green); composite gradational transition zone between CAGr and GrGr (light green grading through to dark green); shear zones (red lines); boreholes (dashed black lines), black numbers adjacent to borehole locations indicate the hydraulic head measured in each borehole (m amsl); the Raterichsboden and Grimsel lake level are at altitudes of 1765 and 1997 m, respectively.

ephemeral streams (Fig. 1a). In addition to soil cover, there are areas of exposed granitoid rock and scree/boulder-covered slopes. Exfoliation fractures, topographical stress fractures and near-vertical tectonic fracture sets cut the surface topography, providing potential surface-water infiltration sites into the groundwater system encompassing the GTS.

Further potential sources of groundwater at the GTS are the surface-water reservoirs. Immediately to the east and south of the GTS, there are two hydropower reservoirs fed by surface runoff and glacial melt (Fig. 1a). These reservoirs are part of a regional pump-storage hydropower network containing multiple reservoirs draining different surface-water catchments. The network of reservoirs is connected by a series of tunnels, pipes and river systems, and reservoir water is regularly pumped both up and down the network, resulting in a well-mixed water body.

Methods

Field sampling

Field sampling took place in August 2018. Surface-soil sites for the sampling of soil organic material were severely restricted by the steep topography of the mountainous slopes above the GTS. Four surface-soil sample sites were selected (sites 1–4 in Fig. 1a) that, as far as possible, describe a north–south transect above the GTS and encompass the visibly different soil/sediment types above the GTS. At each site, two samples were collected *c.* 10 m apart (labelled a and b) to examine the variability of the organic signature at each location. Two aliquots (i and ii) of each soil sample (a and b) from each location (1–4) were subsequently extracted and analysed for their organic signature (see the following 'Sample preparation' and 'Organic analysis' subsections). Soil samples displayed clear differences based on a visual inspection of the soil. Locations 1 and 2 also contained some visible differences between the duplicate samples taken *c.* 10 m apart. Figure 2 shows images of the flora and fauna at each location, and circular markers identify the approximate location where each sample was extracted. Soil at site 1 was dark, waterlogged and had very little sand/gravel content. Soil at site 2 was brown, not as waterlogged as site 1 and contained fragments of roots/plant matter. Soil at site 3 consisted mostly of granite particles and was light brown in colour. Site 4, an ephemeral stream bed, mostly contained angular granitoid rock fragments and fine rock flour.

Lake water was sampled (Fig. 1a) at the one location where the water was safely accessible, and where the predominant SW–NE fracture set within the GTS (Fig. 1b) might plausibly intersect the lake. Due to the highly connected nature of the pump-storage hydropower system (which pumps water between the higher and lower reservoirs), any sample is likely to represent an integrated mixture of surface runoff and glacial meltwaters from both upstream and downstream in the hydropower network. Surface soils were collected using clean metal trowels and placed into aluminium foil parcels. Approximately 1 kg of soil was collected per sample, and the samples were double wrapped and stored at 4°C for transport (below ambient soil temperature at the time of sampling) until sample preparation could take place.

Groundwater was sampled from several locations within the GTS (D, E, F, G, H, I and M in Fig. 1b) from four horizontal boreholes labelled B1–B4 (Fig. 1b) and one vertical borehole B5 drilled from the subsurface gallery upwards towards the surface. All boreholes are fitted with isolated packer systems, integrated with water-sampling flow lines. Volumes of groundwater were sampled from seven individually packed intervals within each borehole (Fig. 1b), which allows for the sampling of water that inflows into a specific section of each borehole. Groundwater sampling intervals (labelled alphanumerically) were chosen to sample the different host-rock lithologies and structural geological features, as well as providing spatial coverage across the GTS; locations D–H are in the Central Aar Granite (CAGr), location I sits in the transition zone between the CAGr and the Grimsel Granodiorite (GrGr), and location M lies in the GrGr. Table 1 describes which sample location (D–M)

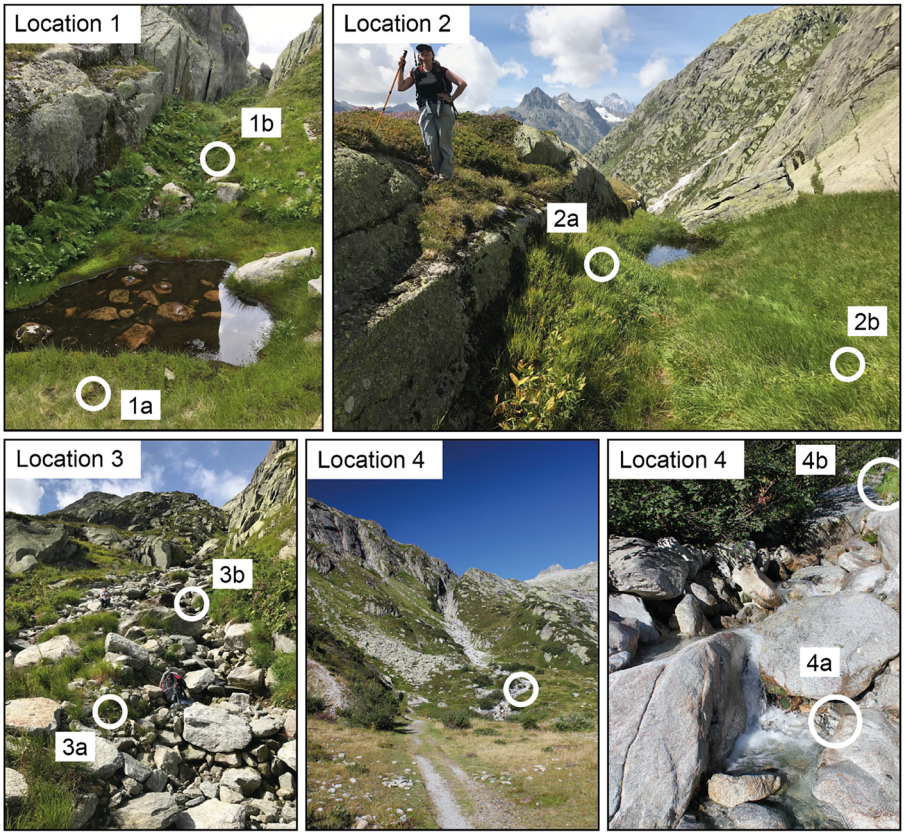

Fig. 2. Location 1 (top left), location 2 (top right), location 3 (bottom left), location 4 overview of river (bottom centre) and specific sample locations of river sediment location 4 (bottom right).

Table 1. The sample types, sample locations, elevation of the sample locations, samples analysed by GC × GC for each sample and for the groundwater sample, the hydraulic head in each sample location (borehole interval), and the borehole the sample location is located within

Sample type	Sample location	Elevation (m amsl)*	Hydraulic head (m amsl)*	Borehole	Samples analysed (GC × GC)†
Groundwater	D	1715	1761	B1	Di, Dii
	E	1702	1825	B1	Ei, Eii
	F	1701	1833	B1	Fi, Fii
	G	1731	1807	B2	G
	H	1731	1912	B3	Hi, Hii
	I	1734	1775	B4	Ii, Iii
	M	1746	1781	B5	M
Lake water	LW	1760			LWi, LWii
Soil/sediment	1	2110			1ai, 1aii, 1bi, 1bii
	2	2140			2ai, 2aii, 2bi, 2bii
	3	1930			3ai, 3aii, 3bi, 3bii
	4	1790			4ai, 4aii, 4bi, 4bii

*m amsl, metres above mean sea level.
†Duplicate groundwater and lake-water samples taken from the same location are denoted by Roman numerals. Duplicate soil samples are denoted by letters a and b, and extraction and analysis of duplicate aliquots of the same soil sample are denoted by Roman numerals.

corresponds to which borehole (B1–B5), the altitude of each sample location and the hydraulic head at each sample location. Groundwater sample locations D, E and F are all located within borehole B1. Prior to groundwater sampling, each borehole interval was drained three times to flush out the volume of the borehole sampling interval itself and the sample lines. Draining of the borehole intervals was carried out to remove any water that had been in contact with plastic in the packer system and to ensure that only the formation water was sampled. Groundwater was used to flush the 125 ml sample bottle three times. Samples were collected and sealed under water with PTFE foil-lined caps (US Geological Survey 2006). Groundwater samples were collected in triplicate for GC × GC and duplicated for CFC analysis; however, during shipping from Switzerland to the UK several samples were smashed or spoiled. As a result, only two samples remained from the lake water, I, F, D, E and H, and only one of the samples taken from G and M were preserved for testing. Thus, we are only able to present the groundwater data analysed by GC × GC either in duplicate for most locations, or as a single sample where only one sample survived. Duplicate samples that have been extracted and analysed by GC × GC are indicated in Roman numerals after the sample location in both the groundwater and lake-water samples (i.e. Di and Dii represent two separate water samples taken from sample location D, extracted and analysed by GC × GC).

Physical and chemical parameters (electrical conductivity, pH, redox potential, dissolved oxygen and temperature) were measured *in situ* during sampling using a flow cell and multiparameter probe (YSI Pro Plus multimeter). Water samples for dissolved ion analysis were filtered using 0.45 µm cellulose acetate filters and stored in 50 ml HDPE centrifuge tubes, with acidified and non-acidified portions stored in a dark fridge and analysed for dissolved ions at the University of Strathclyde within 14 days of sampling, with the exception of alkalinity which was measured on the day of sampling using a HACH digital alkalinity titrator. The methodology used for the sampling and testing of water samples for dissolved inorganic chemistry in this study is described in Stillings *et al.* (2021).

Water samples for 2D gas chromatography (GC × GC) and chlorofluorocarbon (CFC) analysis were collected in 125 ml Boston rounds (borosilicate glass) with foil cap liners. CFC samples were taken according to IAEA (2006) glass bottle collection method 2, and were analysed for CFC-11 and CFC-12 by the British Geological Survey (BGS). The infiltration date and apparent groundwater age were then calculated based on a piston flow model (IAEA 2006) as previously used in the calculation of tritium ages at the GTS (Keppler 1996; Schneeberger *et al.* 2017).

Sample preparation

Each whole soil sample (*c.* 1 kg) was freeze dried and homogenized using a pestle and mortar (washed three times with acetone and a further three times with dichloromethane), then extracted with dichloromethane (DCM) : methanol (MeOH) (9 : 1, v : v) using Accelerated Solvent Extractor (ASE) 350 (Dionex) (US EPA Method 3545A (SW-846): US EPA 2017). ASE extraction cells were packed with 2 g of sample and filled with clean sand heated to 550°C for 8 h. Ground and lake-water samples were extracted by separatory funnel liquid–liquid extraction, using the US EPA Method 3510C as a guideline (US EPA 1996). It was not feasible to transport large volumes of groundwater, so extraction was carried out on a reduced 125 ml sample compared to the EPA method but used the same liquid : liquid ratio. Extraction was carried out three times on each water sample, with a solvent mixture of DCM : MeOH (9 : 1, v : v), to recover the extractable dissolved organic signature from each water sample. DCM was used as the main extraction solvent due to its immiscibility with water and ability to dissolve a wide range of organic compounds that can be detected by electron impact mass spectrometry. The extraction resulted in a wide range of detectible compounds from both the solid and aqueous samples. While this extraction is not exhaustive, it was suitable to construct comparable organic fingerprints of the water and soil samples. Due to the dilute nature of the dissolved organics in groundwater, extracted samples were concentrated using a combination of heat and vacuum concentration (Buchi Syncore Analyst, DCM method) to 1.0 ml volume. Where further sample concentration was required, solvent evaporation with a constant stream of pure N_2 was used to reduce the sample to the desired volume.

Organic analysis

Two-dimensional gas chromatography time of flight mass spectrometry (GC × GC-ToF-MS) operates in a similar way to standard gas chromatography mass spectrometry systems (GC-MS), except that at the end of the first column the compounds are reinjected onto a second column by use of a thermal modulator. This leads to better analyte separation and greater peak intensity. While, in a standard GC-MS, an individual chromatogram peak may represent several different co-eluting compounds, in GC × GC-ToF-MS the co-eluting compounds are more easily separated and identified by their 2D retention times and mass spectra. The following GC × GC-ToF-MS method was used to analyse soil and water extracts, as adapted from the LECO application note (LECO Corporation USA 2019).

Comprehensive signatures of the samples were collected using a LECO (Saint Joseph, MI, USA) time of flight mass spectrometer (Pegasus 4D), with an Agilent 7890A gas chromatography equipped with a LECO thermal modulator. The column set-up was reverse phase, first dimension column DB-17MS (60 m × 0.25 mm i.d. × 0.25 µm: Agilent) polar phase, second dimension column less polar phase Rxi-5Sil MS (1.4 m × 0.25 mm i.d. × 0.25 µm: Restek). Sample injection was splitless using a split/splitless injector set at 260°C, with a helium flow rate of 1.4 ml min^{-1} for the entirety of the run. The primary oven temperature programme was as follows: initial temperature 50°C, hold for 0.2 min, ramp 3.5°C min^{-1} to 320°C, hold for 20 min. The secondary oven and thermal modulator had an offset of +10 and +20°C, respectively, from the primary oven. The thermal modulator period was 5 s, and the mass spectrometer transfer line temperature was 300°C with a spectra acquisition rate of 200 spectra s^{-1}. The instrument method was refined through an iterative process changing the temperature ramp and modulation period until a good peak shape and peak separation were found within a standard compound mix containing a semi-volatile standard with 76 different compounds (8270 Standard, Restek) and a 16 compound C10–C40 (even) *n*-alkane standard (Restek). Repeated injections of the standard compound mix were compared to ensure that the first and second dimension retention times of the compound peaks were consistent across runs. The same standard mix was added as a sample in every subsequent run to ensure that the instrument was working correctly and that the retention times of peaks did not drift between sample batches. This method gave repeatable analyte separation and produced a sufficient number of detectible peaks to build a signature of all the organic compounds contained within the chromatograph.

Data processing and statistical analysis

Processing of 2D gas chromatography data to identify peaks and compounds was carried out using LECO ChromaTOF software. Processing was carried out twice using a low and high signal-to-noise ratio of 50 and 100, respectively. A classification method was applied to remove any compounds related to column bleed or the sample solvent (DCM). The end result for each sample was a 2D chromatograph, and peak table that contained the retention time, intensity, mass spectra and the NIST Library (Linstrom and Mallard 2018) database match for each peak.

The peak tables output from GC × GC analysis can have in excess of 5000 analytes of interest (peaks). Hence, an automated peak table alignment process is required to determine whether peaks with close retention times are genuinely different compounds or whether they are merely misaligned due to a minor shift in retention time between samples. Compound peaks were compared using the statistical compare function within ChromaTOF to align, through pairwise comparison, all the detected peaks within each sample, thus producing an alignment table.

Similarities in the organic fingerprints of each sample were visually identified using PCA (Jolliffe 2002). PCA is a standard technique employed in the analysis of GC × GC data (McGregor *et al.* 2012). PCA of the alignment table was carried out using R (R Team 2018). PCA determines a set of orthogonal axes, or components (linear combinations of the relative concentrations of the organic compounds), that explain the greatest variance within the data using the fewest components. The underlying similarity between samples can then be elucidated by displaying the samples as coordinates of the first two, most explanatory, principal components. Samples that plot at similar locations will contain similar combinations (or patterns) of the organic compounds. To determine which samples are most similar, hierarchical cluster analysis (HCA) was also performed using R (R Team 2018). HCA determines the shortest distance between each sample. HCA was carried out using the H-clust function within R (R Team 2018). HCA algorithms continue to pair the closest samples, based on their Euclidean distance, until the whole dataset is described within the same cluster. HCA finds a series of clusters that identify similarities between samples. To identify whether clusters were statistically significant, a *P*-value was calculated using the multiscale bootstrapping 'pvclust' function in R and adopting the approximately unbiased approach, as described in Suzuki and Shimodaira (2006).

Results

The groundwater at the GTS is of low conductivity (69–84 µS cm^{-1}) and alkaline (pH 8.83–9.39). Borehole intervals in the south of the GTS have higher dissolved sodium and lower dissolved calcium concentrations than borehole intervals in the north of the GTS (Fig. 3a; see also the data in Supplementary Table S1); these results are consistent with previous findings (Schneeberger *et al.* 2017; Stillings *et al.* 2021) that have been shown by Schneeberger *et al.* (2017) to reflect the change in host-rock lithology from Grimsel Granodiorite (GrGr) in the south of the GTS to Central Aar Granite (CAGr) in the north of the GTS. Groundwater residence time estimates vary between sampling locations in the GTS, which is likely to reflect poor connectivity in the fracture network between surface recharge and sampling locations at depth, giving rise to variably tortuous flow paths (Stillings *et al.* 2021). The CFC concentrations in samples taken from intervals F, G, H and I, according to IAEA (2006) method 2, and analysed for CFC-11 and CFC-12 by the BGS indicated an apparent groundwater residence time of 57–67 years at this location based on a piston flow model (IAEA 2006). The apparent groundwater age from CFC measurements is consistent with historical and recent tritium measurements. Early tritium measurements from two boreholes (not sampled here) in the south of the GTS (Keppler 1996; Schneeberger *et al.* 2019) showed an apparent groundwater residence time of between 5 and 36.5 years. More recently, tritium measurements from interval G (see Supplementary Table S1) imply an apparent groundwater age of more than 60 years (Schneeberger *et al.* 2017). $^{14}C_{DOC}$ dating (Keppler 1996; Schneeberger *et al.* 2019) shows a similar difference in age estimates, with apparent residence times of 220 ± 180 years for intervals D, E and F in the north, and 13 ± 3 years for intervals (not sampled here) in the south.

In general, the results of the GC × GC analysis show that the samples are highly complex and contain a large number of organic compounds. By way of example, typical GC × GC chromatograms, in the form of 2D contour plots, for soil sample 2b and groundwater sample G are shown in Figure 3b and c. The colour temperature scale denotes high-intensity areas, and each high point represents an individual compound peak. The same compound in each sample will occupy approximately the same retention time in both the first (*x*-axis) and second dimensions (*y*-axis), and will plot at the same location on each chromatograph. Similar compounds or groups of compounds elute along predictable trends in the chromatograph. Changes in chain length of the same type of compounds (i.e. carbon number) are reflected by a systematic increase in the retention time. Different groups of compounds have different affinities to the second-dimension column's stationary phase, causing separation by the compound group along the *y*-axis. So, for example, *n*-alkanes have a clear peak separation from ketones, alcohols, aromatics, and other branched and unsaturated aliphatic compounds. When comparing the soil sample to the groundwater sample (Fig. 3b, c), there are similar patterns in the elution of specific compounds. However, the relative concentration of the longer chain alkanes and alkenes (labelled in Fig. 3b) is higher in the soil than in the groundwater. The total number of compounds detected in each

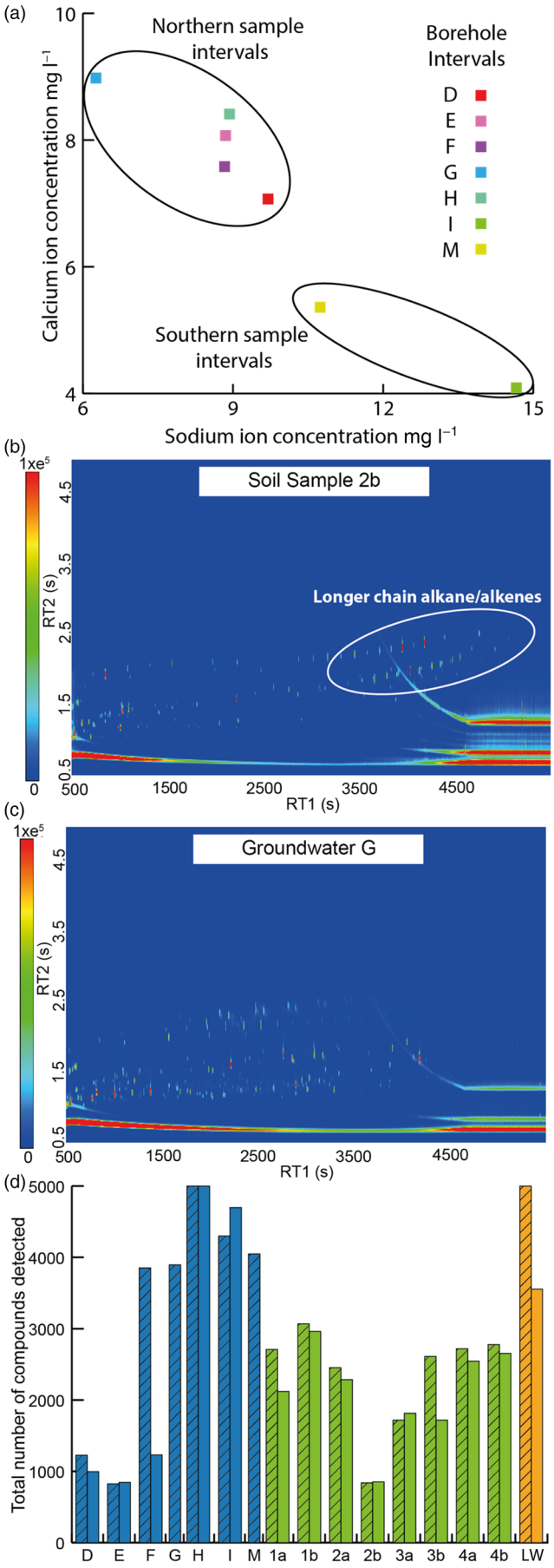

Fig. 3. (**a**) Average ($n = 26$) groundwater calcium v. sodium ion concentration (mg l^{-1}). (**b**) and (**c**) GC × GC 2D chromatographs of (b) soil sample 2b and (c) groundwater G. RT1 represents the first dimension retention time in seconds and thus separation in column 1, RT2 represents the second dimension retention time in seconds and compound separation in column 2. The colour scale reflects the total ion count (TIC), where red reflects a high TIC and blue a low TIC; red/light blue dots represent individual compounds; the region of long-chain alkanes and alkenes is indicated with a white circle in (b). Long coloured streaks parallel to the *x*-axis in the bottom of the chromatograph are artefacts of the sample matrix and GC column stationary phase, and are not included in any data analysis. (**d**) The total number of organic compounds detected in each sample: the first replicate sample (i) from the same location is indicated by diagonal lines across the colour and the second replicate sample (ii) is indicated by a solid colour.Source: (a) methodology taken from Stillings *et al.* (2021).

sample is given in Figure 3d; compound abundance varies significantly, ranging from 826 in groundwater sample E to 5000 in lake-water sample LW.

To identify similarities in the organic signatures of surface and groundwater samples, it is necessary to examine the relative abundance of individual compounds that are common to most samples: that is, to determine whether the ratios (or pattern) of preserved compounds at depth can be compared with the surface-soil and water samples, and, hence, used to indicate the groundwater origin. The statistical compare function within ChromaTOF was used to align the organic compounds, producing a compound alignment table (see Supplementary Table S2). The statistical analysis used 50 organic compounds, which were common to 80% of the samples. To calculate the relative abundance of each compound, the abundance ratios of these 50 aligned compounds within an individual sample was taken (i.e. for each sample the sum of the relative concentrations of all 50 compounds is equal to 1). Figure 4 summarizes the relative abundance of compound classes for the 50 aligned compounds for each sample, where the length of each colour represents the relative fraction of each compound classification. Most repeat samples from the same sample location (labelled i and ii) have similar proportions of each different compound class, and samples of the same type (i.e. groundwater, soil and lake water) are visibly similar (Fig. 4) – with the exceptions of D and E from the groundwater, and 2b from the soil, all three of which have a smaller proportion of acid and alcohol compounds than the other samples. There is a visible variation in the relative abundance compound classes within the groundwater samples F, G, H, I and M that most notably vary in their organic acid content, where the relative abundance ranges from 20 to 50% of the aligned compounds.

PCA and HCA results

Before comparing organic fingerprints between surface samples and groundwater samples, it is first important to understand the variability of aligned compounds in the surface samples to determine if there is a variation in soil organic fingerprint that may be reflected in the groundwater samples. Comparison of all the aligned soil samples using the 50 aligned organic compounds common to all samples as explanatory variables was carried out using PCA (Fig. 5a), and clustering was performed using HCA to identify any significant clusters (Fig. 5b). In both the PCA and the HCA the replicate samples (i and ii) plot in the same location, with the exception of samples 1ai and 1aii that do not plot as closely. Cluster analysis shows that there is more than 95% confidence of two different clusters being present within the soil samples. Locations 1–4 cluster together, while 2b clusters separately from 2a and the other soil sample locations.

To identify whether similarities exist between the groundwater samples, the lake water and the different soil clusters identified in Figure 5b, PCA was carried out using the same 50 aligned organic compounds as explanatory variables, as determined from the peak alignment table (see Supplementary Table S2). The analysis used all

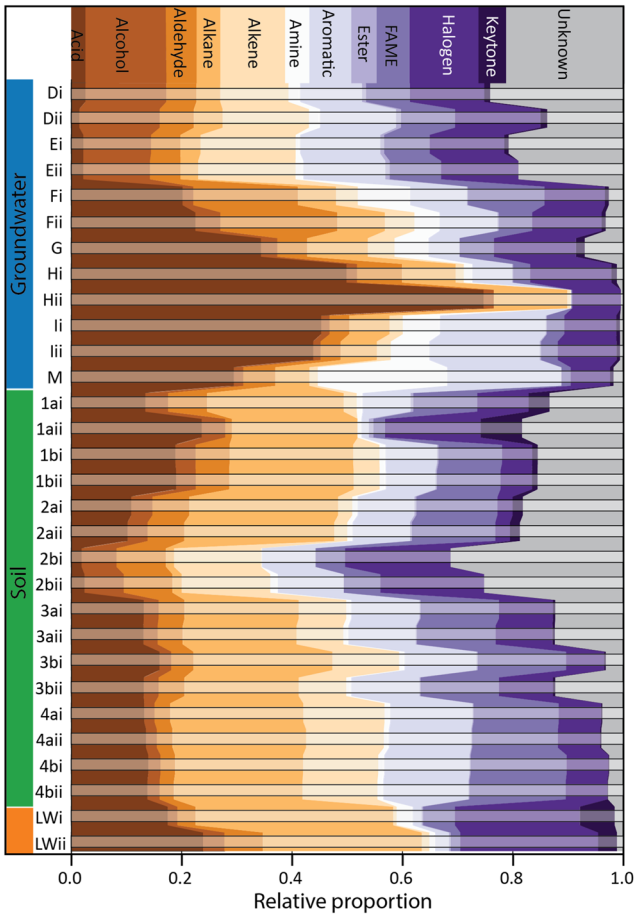

Fig. 4. Concentration-normalized bar plots showing the relative ratio of each compound classification present in the aligned dataset of all 50 aligned compounds for each sample. Each horizontal bar represents the relative change in compound classes within each duplicate sample. Samples are separated into groundwater, soil and lake water (LWi, LWii).

samples (soil, lake water and groundwater) in order to determine whether the surface and groundwater sampling sites have clearly distinct organic signatures that form statistically significant clusters based on the aligned compounds between samples. Results from the PCA are shown in Figure 6. Principal component 1 (PC1) explains 33% of the variance and principal component 2 (PC2) explains 19% of the variance, which is not uncommon in datasets such as this with a larger number of variables compared to the number of observations (Ringnér 2008). Most repeat soil and sediment samples (labelled i and ii for direct repeats, and a and b for samples taken in the same geographical area) plot in the same region of the PCA plot (Fig. 6). Soil samples from sites 3 and 4, taken from the lower slopes of the mountain above the GTS, consistently plot in a similar location. Of the two samples taken from location 2, sample 2a plots similarly to the other soils; however, sample 2b, whilst being similar in PC1, is very different in PC2 when compared to all of the other soil samples. This indicates a distinct difference in the organic signature of soils at location 2, which is at the highest elevation of all of the soil sampling sites and was the only location at which roots and plant matter were visually apparent in the soil samples. The lake water (orange) plots in a distinctly different location to any of the soil samples (green), indicating it has a different organic signature that is clearly distinguished within PC1.

Groundwater samples F, G, H, I and M all plot close to the lake-water signature in the PCA (Fig. 6), indicating that they have similar organic fingerprints to each other and could be derived predominantly from infiltration of the lake water. The water level in Lake Raterichsboden lies within the range of groundwater head measurements found throughout the GTS (at the time of sampling the water level in Raterichsboden was higher than the head in interval I but lower than in intervals F, G and M). By comparison, the water level in Lake Grimsel is higher than all head measurements in the GTS. Hence, it may be that the groundwaters in intervals F, G, H, I and M were originally derived from Lake Grimsel, which is immediately upstream of Lake Raterichsboden and hydraulically connected via the pumped-storage hydropower system.

Groundwater samples D and E differ from the other groundwater samples (blue in Fig. 6). All groundwater and lake-water samples contain 44 or more of the 50 aligned compounds (present in 80% of samples), with the exception of Hii, which contains 21 out of the 50 and plots near to Hi; so the differences in PC1 and PC2 for samples D and E cannot be attributed to a smaller number of compounds in these samples. Samples D and E plot with soil sample 2b (Fig. 6), indicating that the groundwaters contain a similar organic signature to soils at higher elevations and are likely to have originated from surface-soil infiltration. Interestingly, groundwater samples D, E and F are all sampled from separate locations at different distances down the borehole; sample locations are separated by a hydraulic packer system with separate flow lines to allow sampling from different distances within the same horizontal borehole. Groundwaters at D and E seem to comprise predominantly water

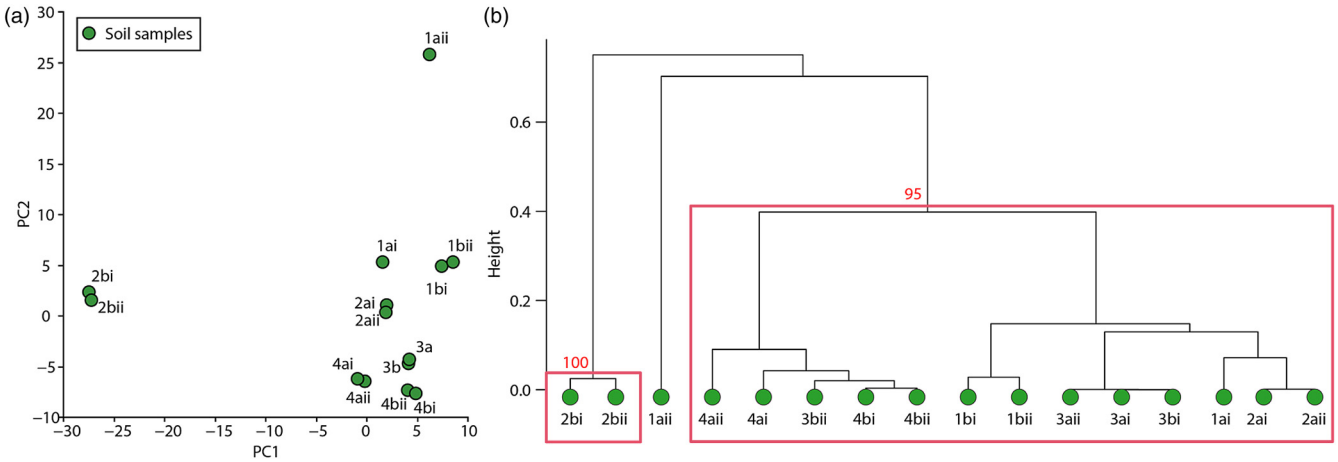

Fig. 5. (**a**) PCA plot of the soil sample locations based on the 50 aligned compounds present in 80% of all samples and (**b**) HCA plot showing the >95% cluster confidence calculated through approximately unbiased bootstrapping (red numbers and box). Replicates for sample locations 3a and 3b are superimposed on top of each other in the PCA in (a).

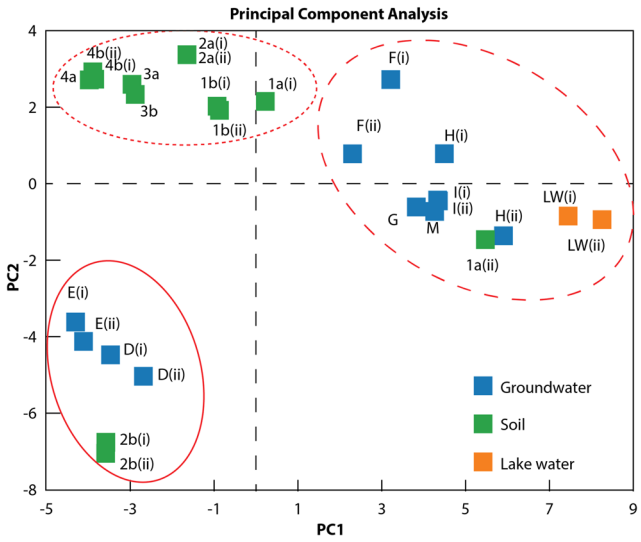

Fig. 6. PCA plot of each sample based on the alignment table of 50 organic compounds. All points are labelled with the sample name. Groundwater samples (blue), soil (green) and lake water 'LW' (orange). Principal component 1 (PC1) is along the *x*-axis and principal component 2 (PC2) is on the *y*-axis. Markers for replicate samples (i and ii) for the sample locations 3a, 3b and 4a are superimposed on top of each other.

that infiltrated through surface soils, whereas groundwater at interval F, in the same borehole, is likely to have a hydraulic connection to the lakes or is derived from lake water. Their very different organic signatures supports previous research observations that the local fracture network is very poorly connected even between sampling intervals from the same borehole (Stillings *et al.* 2021).

To identify whether the clusters indicated by the PCA analysis (Fig. 6) are statistically significant, HCA was carried out on the 50 aligned compounds (Fig. 7). Three statistically significant (99% confidence level) clusters are identified and grouped through HCA, indicating that the samples within each cluster have statistically similar organic fingerprints. Within each cluster on the dendrogram (Fig. 7), as expected, the series neighbour (the most similar) in the HCA analysis are replicate samples collected from the same borehole interval or, in the case of the soils, the duplicate extraction of the same homogenized soil sample (i.e. i and ii). In the soils, the next closest pair is generally the 'b' sample from a neighbouring location, except for soil samples 1a and 3b. The only sample that does not cluster with its duplicate is sample 1aii, which clusters with the groundwaters, while all the other samples taken from location 1 cluster together within the other soil samples 2a, 3a, 3b, 4a and 4b; we attribute the deviation of sample 1aii to analytical uncertainty probably during the extraction of the organic fingerprint from sample 1aii, as a result this point can be considered as an outlier. For all 12 separate samples taken from the eight groundwater sampling intervals (Figs 6 and 7), the HCA analysis shows it is possible to group and differentiate samples with similar organic fingerprints, which implies that there is likely to be a relationship between samples within the same clusters.

The top 10 compounds with the highest loading magnitudes in PC1 are the compounds that describe the most sample-to-sample variance in the first principal component, presented in Table 1, and, hence, are indicative of the differences between clusters in the organic fingerprints. PC1 is responsible for the separation of the 'groundwater and lake-water cluster' from the other two clusters shown in Figure 6. Of the top 10 loading compounds in PC1 some derive from natural sources, for example: decane, 6-4ethyl-2-methyl- has been found in a certain species of plant root extract (Shettima *et al.* 2013); 1-iodoundecane is commonly found in mammal urine (Achiraman and Archunan 2002) and is also an active compound in some plants (Khammas *et al.* 2020); and 1-hexene, 4,5-dimethyl- has been detected as an excretory compound from fungi (Simon *et al.* 2017). Of the top 10 compounds in PC2, which separates the smaller 'groundwater and soils' cluster from the other two clusters in Figure 6, two are classified as unknown compounds (unknown compound 147 and 224) as they do not have a higher than 70% match to any compounds within the NIST Library (Linstrom and Mallard 2018). A manual search of unknown 147 and unknown 224 with the library found a 66% match with propanoic acid, 3-hydroxy-2-isopropylidene and a 63% match with 1,1-difluoro-2,2-dimethyl-cyclopropane, respectively. Two other compounds in the top 10 loadings in PC2 are known to derive from natural sources: oxime-, methoxy-phenyl- has been found in plants and in animal mucus (Sallam *et al.* 2009; Al-Mussawii *et al.* 2022); 2-undecenal, E- is found in essential oils derived from some plants (Kivcak *et al.* 2001); and decane has also been found in plants (Cakir *et al.* 2004) but is also common in petroleum and coal tars (Pan *et al.* 2012). Thus, decane does not have a specifically discernable natural source. Other compounds (PC1 and PC2 in Table 2) do not have a specific identifiable natural source and could potentially derive from either natural or industrial processes.

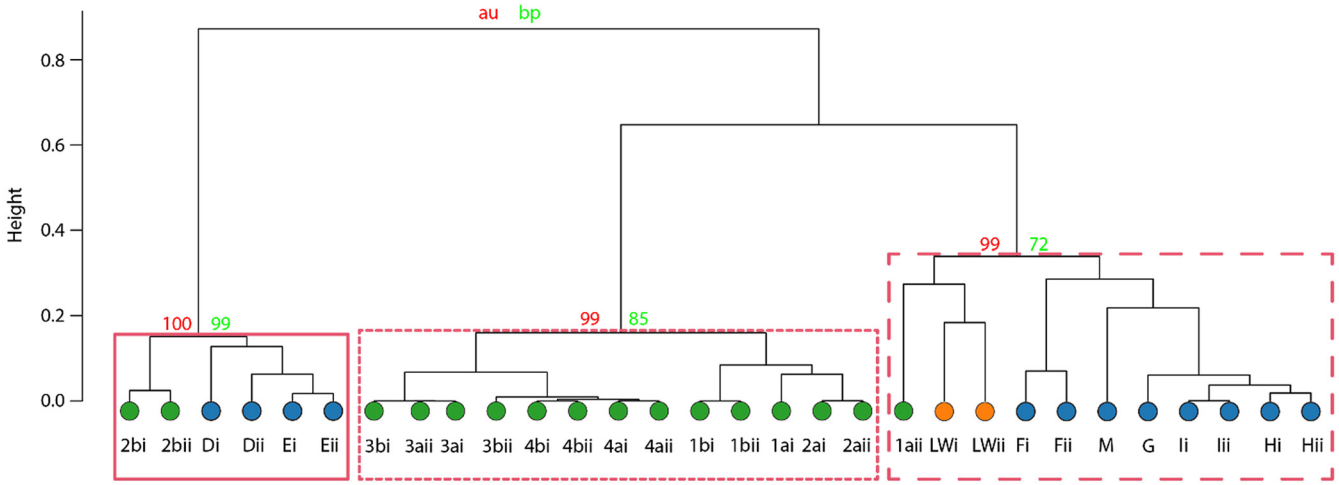

Fig. 7. Dendrogram based on the HCA of the groundwater, lake-water, soil and sediment samples. Groundwater samples (blue) are similar to lake water (orange) and samples similar to soils/sediment (green). Bootstrapping is performed using an approximately unbiased (AU) approach, which is calculated using multiple scale bootstrap resampling and bootstrap probability (BP) that is performed through standard bootstrap resampling at the same scale.

Table 2. Top 10 highest loading magnitude compounds in PC1 and PC2 from the PCA based on the 50 aligned compounds present in 80% of the samples

Highest loading magnitudes PC1	Highest loading magnitudes PC2
1-Heptene, 5-methyl-	Unknown compound 147*
Benzene, 1,2,3-trimethyl-	Benzene, 1-ethyl-2,4-dimethyl-
Decane, 6-ethyl-2-methyl-	Oxime-, methoxy-phenyl-
1-Iodoundecane	2-Undecenal, E-
1-Hexene, 4,5-dimethyl-	Unknown compound 224*
Tetradecane, 1-iodo-	Decane
Acetic acid, butyl ester	Benzene, 1-methyl-3-propyl-
Cyclohexane, 1-isopropyl-1-methyl-	Benzene, 1-ethyl-4-methyl-
1-Iodoundecane	2-Decenal, (Z)-
2-(3-Hydroxy-2-nitrocyclohexyl)-1-phenylethanone	Phosphonic acid, (*p*-hydroxyphenyl)-

*Compounds with less than 70% match to the NIST Library.

Discussion

The results of the PCA and HCA analyses show that groundwater, lake water and soil waters have distinct, and repeatable (i.e. duplicate samples that fall within the same cluster), dissolved organic signatures, and that these signatures can be potentially be used to determine the predominant influence of near-surface recharge sources on groundwater samples at depth. The PCA and HCA analysis identified three statistically significant clusters, based on their organic signatures. These clusters indicate that at the GTS, most groundwater sampling intervals tap fractures that are hydraulically connected to the lake water. However, sampling intervals D and E, which are in the north of the GTS do not cluster with the lake water, this suggests a second potential infiltration source that could reflect water infiltrating through soils at higher altitudes and which have an organic signature similar to soil sample 2b. The variability in the organic groundwater signatures, particularly from neighbouring sampling intervals within the same borehole (D, E and F), underlines the poorly connected nature of the fracture network. This observation is further supported by the variation in the groundwater age estimates between boreholes at different locations (Keppler 1996; Schneeberger *et al.* 2019) and previous observations of highly localized perturbations in pH associated with microseismic events during reservoir drainage and maintenance (Stillings *et al.* 2021).

Previously researchers have successfully discriminated groundwater origins through the use of unique 'target' biomarker compounds (Derrien *et al.* 2017). At the GTS, no such target biomarker compounds were found. Instead, groundwater origins were obtained using an untargeted organic 'fingerprint' for each water/soil sample, in which relative concentrations were determined for a large number of common compounds. For organic fingerprinting to be an effective groundwater tracer at other locations, a sufficient number of organic compounds within the surface signatures must be well preserved over time, so as to be identifiable at depth. Organic matter decay rates change most rapidly in the shallow subsurface (to depths of *c*. 350 m), attributed to the changes in oxidative decomposition (Kortelainen and Karhu 2006). However, active microbial communities have been shown to exist in groundwater systems to depths of up to *c*. 1 km (e.g. Shimizu *et al.* 2007; Nyyssönen *et al.* 2012). These communities will gradually metabolize organic components in the groundwater, leading to progressive degradation of organic parent molecules over time and, hence, along the groundwater flow paths. This degradation of organic parent molecules, into daughter decay products, is likely to explain the large total number of compounds (Fig. 3d) that were found in the lake water and most groundwater samples when compared with the soil samples. At the GTS, where apparent groundwater residence times vary by location, from 5 to 220 years (Keppler 1996; Schneeberger *et al.* 2017, 2019), the number of detectable organic compounds that remained in 80% of all samples, and were thus usable in the final PCA and HCA analysis, was 50. Our results show that these 50 compounds were sufficient to discriminate between the distinct surface environments, and that these surface organic signatures could still be identified in groundwaters at depth. Evidence from other research fields also suggests that long-term solid and dissolved preservation of organic compounds may not be uncommon; Korkmaz and Gülbay (2007) used specific compounds as indicators of surface deposition environments for petroleum source rocks that are of Jurassic age (Korkmaz and Gülbay 2007), while specific dissolved organic compounds are found preserved in groundwaters of up to 23 kyr in age (Aravena *et al.* 1995). Further studies, using sites with older and younger groundwaters in differing surface and geological environments, are required to determine the range of geological settings and age of groundwaters for which organic fingerprinting can prove a useful tool for investigating groundwater origins.

It is possible that some of the variability in the PCA analysis in Figure 6 is due to groundwater mixing between surface infiltration water and the lake water, specifically at groundwater sampling location F. Whilst F forms a statistically significant cluster (at >99% confidence level) with the lake water in the HCA analysis, in the PCA it plots between some of the soils and the lake water. To identify whether groundwater mixing could be responsible, future investigations could prepare different proportional mixes of each infiltration source and include these signatures for comparison in the statistical analysis, thereby allowing any potential groundwater mixing to be identified.

The conclusions that can be drawn about the groundwater system at the GTS in this study are limited due to the small total number of samples that it was possible to collect and the restrictions to surface access. Despite the small sample size, the HCA identified three clusters with a 99% confidence level, enabling us to clearly distinguish groundwater sampling locations that are predominantly fed via surface-soil infiltration (D and E) from those that are dominated by lake-water infiltration. In future studies, a larger sample size would reduce the uncertainty when comparing organic fingerprints and might enable clusters to be identified that link individual surface-soil infiltration sites to specific groundwater sampling locations. The use of additional complementary geochemical techniques would also be useful in further constraining the meteoric infiltration locations.

Summary

Two-dimensional gas chromatography was used to organically fingerprint surface-soil, lake-water and groundwater samples at the Grimsel Test Site in Switzerland. Three distinct meteoric infiltration types were identified with uniquely different proportions of the same compounds forming their individual organic fingerprints: two types of surface-soil environment and the lake water. These surface infiltration fingerprints were compared to organic signatures found within seven borehole sampling intervals located at a depth of 200–500 m below ground surface, positioned throughout the length of the GTS tunnels. Fifty organic molecules were found to be common to 80% of samples. Using principal component analysis (PCA), the relative abundance of these molecules was used to match the individual borehole samples to their likely surface-water origins. Hierarchical cluster analysis (HCA) was used to identify three statistically significant clusters. These clusters showed that most groundwater sampling intervals were clustered with the lake water and, hence, primarily tap fractures that are hydraulically connected to the lake. Two intervals, however, were clustered with soil taken from the highest altitude sampled on the mountain above the GTS,

thus suggesting that these tap fractures connected to surface infiltration water at high altitudes. This research demonstrates that natural organic molecules, and their relative abundance, are sufficiently well preserved in groundwater over timescales of several decades that they can be used to discriminate the near-surface environment(s) through which meteoric groundwater has infiltrated. Organic fingerprinting could be a powerful new tool for an improved understanding of groundwater flow systems, particularly when used in combination with other complementary tracing techniques.

Acknowledgements The research forms part of the collaborative Large-Scale Monitoring (LASMO) programme at the Grimsel Test Site. We would like to thank the anonymous reviewers for their suggestions and helpful input.

Author contributions **MS**: conceptualization (supporting), data curation (lead), formal analysis (equal), investigation (lead), methodology (lead), validation (equal), visualization (equal), writing – original draft (lead), writing – review & editing (lead); **RJL**: conceptualization (equal), formal analysis (supporting), funding acquisition (lead), project administration (lead), supervision (equal), visualization (supporting), writing – original draft (supporting), writing – review & editing (equal); **ZKS**: conceptualization (equal), funding acquisition (equal), investigation (equal), project administration (equal), supervision (equal), visualization (supporting), writing – original draft (supporting), writing – review & editing (equal); **RAL**: conceptualization (equal), funding acquisition (equal), investigation (equal), methodology (equal), project administration (equal), validation (equal), writing – review & editing (equal); **ST**: conceptualization (equal), funding acquisition (equal), project administration (equal), writing – review & editing (equal); **MK**: formal analysis (supporting), methodology (equal), project administration (supporting), resources (equal), software (equal), validation (equal).

Funding This work was funded by the Engineering and Physical Sciences Research Council (grant EP/M506643/1 awarded to R.A. Lord) and the Nuclear Waste Services (grant awarded to R.J. Lunn).

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability Data that support the figures are available in the Supplementary information dataset ds01.

References

Abrantes, J.R., Moruzzi, R.B., Silveira, A. and de Lima, J.L. 2018. Comparison of thermal, salt and dye tracing to estimate shallow flow velocities: Novel triple-tracer approach. *Journal of Hydrology*, **557**, 362–377, https://doi.org/10.1016/j.jhydrol.2017.12.048

Achiraman, S. and Archunan, G. 2002. Characterization of urinary volatiles in Swiss male mice (*Mus musculus*): bioassay of identified compounds. *Journal of Biosciences*, **27**, 679–686, https://doi.org/10.1007/BF02708376

Almstetter, M.F., Oefner, P.J. and Dettmer, K. 2012. Comprehensive two-dimensional gas chromatography in metabolomics. *Analytical and Bioanalytical Chemistry*, **402**, 1993–2013, https://doi.org/10.1007/s00216-011-5630-y

Al-Mussawii, M.A., AL-Sultan, E.Y., AL-Hamdani, M.A. and Ramadhan, U.H. 2022. Antibacterial activity of alkaloid compound Methoxy phenyl–Oxime ($C_8H_9N0_2$) isolated and purified from leaf of *Conocarpus lancifolius* Engl. *Teikyo Medical Journal*, **45**, 4971–4981.

Amaral, M.S.S., Nolvachai, Y. and Marriott, P.J. 2020. Comprehensive two-dimensional gas chromatography advances in technology and applications: biennial update, *Analytical Chemistry*, **92**, 85–104, https://doi.org/10.1021/acs.analchem.9b05412

Aravena, R., Wassenaar, L.I. and Plummer, L.N. 1995. Estimating ^{14}C groundwater ages in a methanogenic aquifer. *Water Resources Research*, **31**, 2307–2317, https://doi.org/10.1029/95WR01271

Avrahamov, N., Yechieli, Y., Purtschert, R., Levy, Y., Sültenfuß, J., Vergnaud, V. and Burg, A. 2018. Characterization of a carbonate karstic aquifer flow system using multiple radioactive noble gases (^3H–^3He, ^{85}Kr, ^{39}Ar) and ^{14}C as environmental tracers. *Geochimica et Cosmochimica Acta*, **242**, 213–232, https://doi.org/10.1016/j.gca.2018.09.009

Baker, M.A., Valett, H.M. and Dahm, C.N. 2000. Organic carbon suppply and metabolism in shallow groundwater ecosystems. *Ecology*, **81**, 3133–3148, https://doi.org/10.1890/0012-9658(2000)081[3133:OCSAMI]2.0.CO;2

Cakir, A., Kordali, S., Zengin, H., Izumi, S. and Hirata, T. 2004. Composition and antifungal activity of essential oils isolated from *Hypericum hyssopifolium* and *Hypericum heterophyllum*. *Flavour and Fragrance Journal*, **19**, 62–68, https://doi.org/10.1002/ffj.1279

Casillas-Trasvina, A., Rogiers, B., Beerten, K., Pärn, J., Wouters, L. and Walraevens, K. 2022. Using helium-4, tritium, carbon-14 and other hydrogeochemical evidence to evaluate the groundwater age distribution: The case of the Neogene aquifer, Belgium. *Journal of Hydrology X*, **17**, 100132, https://doi.org/10.1016/j.hydroa.2022.100132

Chen, M., Price, R.M., Yamashita, Y. and Jaffé, R. 2010. Comparative study of dissolved organic matter from groundwater and surface water in the Florida coastal Everglades using multi-dimensional spectrofluorometry combined with multivariate statistics. *Applied Geochemistry*, **25**, 872–880, https://doi.org/10.1016/J.APGEOCHEM.2010.03.005

Derrien, M., Yang, L. and Hur, J. 2017. Lipid biomarkers and spectroscopic indices for identifying organic matter sources in aquatic environments: A review. *Water Research*, **112**, 58–71, https://doi.org/10.1016/j.watres.2017.01.023

Fackrell, J.K., Glenn, C.R., Thomas, D., Whittier, R. and Popp, B.N. 2020. Stable isotopes of precipitation and groundwater provide new insight into groundwater recharge and flow in a structurally complex hydrogeologic system: West Hawai'i, USA. *Hydrogeology*, **28**, 1191–1207, https://doi.org/10.1007/s10040-020-02143-9

Filippini, M., Squarzoni, G. *et al.* 2018. Differentiated spring behavior under changing hydrological conditions in an alpine karst aquifer. *Journal of Hydrology*, **556**, 572–584, https://doi.org/10.1016/j.jhydrol.2017.11.040

Flury, M. and Wai, N.N. 2003. Dyes as tracers for vadose zone hydrology. *Reviews of Geophysics*, **41**, 102, https://doi.org/10.1029/2001RG000109

Gerber, C., Vaikmäe, R. *et al.* 2017. Using ^{81}Kr and noble gases to characterize and date groundwater and brines in the Baltic Artesian Basin on the one-million-year timescale. *Geochimica et Cosmochimica Acta*, **205**, 187–210, https://doi.org/10.1016/j.gca.2017.01.033

IAEA 2006. *Use of Chlorofluorocarbons in Hydrology: A Guidebook*. STI/PUB/1238. International Atomic Energy Agency (IAEA), Vienna.

Jolliffe, I.T. 2002. Intoduction. *In*: *Principal Component Analysis*. 2nd edn. Springer Series in Statistics. Springer, New York, 1–9, https://doi.org/10.1007/0-387-22440-8_1

Keiluweit, M., Nico, P.S., Kleber, M. and Fendorf, S. 2016. Are oxygen limitations under recognized regulators of organic carbon turnover in upland soils? *Biogeochemistry*, **127**, 157–171, https://doi.org/10.1007/s10533-015-0180-6

Keppler, A. 1996. *Hydrogeologische, Hydrochemische und Isotopenhydrologische Untersuchungen an den Oberflächen- und Kluftwässern im Grimselgebiet, Schweiz*. GSF Bericht, **4/96**.

Khammas, A.D., Abd Ali, N. and Amin, M.K.M. 2020. Diagnosis of active compounds of ethanol and hexane extract of *Azadirachta indica* by gas chromatograph mass spectrometer technique (GC-MS) and evaluation of its antagonism efficiency on the *Fusarium oxysporum* f. sp. *lycopersici* growth *in vitro*. *Plant Archives*, **20**, 6741–6750.

Khatri, N. and Tyagi, S. 2015. Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Frontiers in Life Science*, **8**, 23–39, https://doi.org/10.1080/21553769.2014.933716

Kivcak, B., Mert, T., Demirci, B. and Baser, K.H.C. 2001. Composition of the essential oil of *Arbutus unedo*. *Chemistry of Natural Compounds*, **37**, 445–446, https://doi.org/10.1023/A:1014419309885

Korkmaz, S. and Gülbay, R.K. 2007. Organic geochemical characteristics and depositional environments of the Jurassic coals in the eastern Taurus of Southern Turkey. *International Journal of Coal Geology*, **70**, 292–304, https://doi.org/10.1016/j.coal.2006.07.001

Kortelainen, N.M. and Karhu, J.A. 2006. Tracing the decomposition of dissolved organic carbon in artificial groundwater recharge using carbon isotope ratios. *Applied Geochemistry*, **21**, 547–562, https://doi.org/10.1016/j.apgeochem.2006.01.004

Kralik, M., Humer, F. *et al.* 2014. Using $^{18}O/^2H$, $^3H/^3He$, ^{85}Kr and CFCs to determine mean residence times and water origin in the Grazer and Leibnitzer Feld groundwater bodies (Austria). *Applied Geochemistry*, **50**, 150–163, https://doi.org/10.1016/j.apgeochem.2014.04.001

LECO Corporation USA 2019. *Application Note. Instrument: Pegasus® BT 4D. Determining Terpene Profiles of Cannabis Strains Using GC and GCxGC with High Performance TOFMS*. LECO Corporation USA, Saint Joseph, MI, https://eu.leco.com/images/SepSci-Application-Library/PEGBT4D_CANNABIS_TERPENE_PROFILES_203-821-558.pdf

Linstrom, P.J. and Mallard, W.G. 2018. *NIST Chemistry WebBook: NIST Standard Reference Database Number 69*. National Institute of Standards and Technology (NIST), Gaithersburg MD, https://doi.org/10.18434/T4D303

McGregor, L.A., Gauchotte-Lindsay, C., Nic Daéid, N., Thomas, R. and Kalin, R.M. 2012. Multivariate statistical methods for the environmental forensic classification of coal tars from former manufactured gas plants. *Environmental Science & Technology*, **46**, 3744–3752, https://doi.org/10.1021/es203708w

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P. and Itävaara, M. 2012. Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in Olkiluoto, Finland. *Geomicrobiology Journal*, **29**, 863–878, https://doi.org/10.1080/01490451.2011.635759

Obernosterer, I. and Benner, R. 2004. Competition between biological and photochemical processes in the mineralization of dissolved organic carbon.

Limnology and Oceanography, **49**, 117–124, https://doi.org/10.4319/lo.2004.49.1.0117

Okuhata, B.K., Thomas, D.M., Dulai, H., Popp, B.N., Lee, J. and El-Kadi, A.I. 2022. Inference of young groundwater ages and modern groundwater proportions using chlorofluorocarbon and tritium/helium-3 tracers from West Hawai ʻi Island. *Journal of Hydrology*, **609**, 127755, https://doi.org/10.1016/j.jhydrol.2022.127755

Osman Awaleh, M., Boschetti, T. *et al.* 2020. Hydrochemistry and multi-isotope study of the waters from Hanlé–Gaggadé grabens (Republic of Djibouti, East African Rift System): A low-enthalpy geothermal resource from a transboundary aquifer. *Geothermics*, **86**, 101805, https://doi.org/10.1016/j.geothermics.2020.101805

Pan, N., Cui, D. *et al.* 2012. Characterization of middle-temperature gasification coal tar. Part 1: bulk properties and molecular compositions of distillates and basic fractions. *Energy & Fuels*, **26**, 5719–5728, https://doi.org/10.1021/ef3007323

Patrushev, Y.V. 2015. Advantages of two-dimensional gas chromatography. *Kinetics and Catalysis*, **56**, 386–393, https://doi.org/10.1134/S0023158415030155

Pichler, T. 2005. Stable and radiogenic isotopes as tracers for the origin, mixing and subsurface history of fluids in submarine shallow-water hydrothermal systems. *Journal of Volcanology and Geothermal Research*, **139**, 211–226, https://doi.org/10.1016/j.jvolgeores.2004.08.007

Prada, S., Virgílio Cruz, J. and Figueira, C. 2016. Using stable isotopes to characterize groundwater recharge sources in the volcanic island of Madeira, Portugal. *Journal of Hydrology*, **536**, 409–425, https://doi.org/10.1016/j.jhydrol.2016.03.009

Ringnér, M. 2008. What is principal component analysis? *Nature Biotechnology*, **26**, 303–304, https://doi.org/10.1038/nbt0308-303

R Team 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, https://www.r-project.org

Sallam, A.A.A., El-Massry, S.A. and Nasr, I.N. 2009. Chemical analysis of mucus from certain land snails under Egyptian conditions. *Archives of Phytopathology and Plant Protection*, **42**, 874–881, https://doi.org/10.1080/03235400701494448

Schneeberger, R., Berger, A., Herwegh, M., Eugster, A., Kober, F., Spillmann, T. and Blechschmidt, I. 2016. *GTS Phase VI – LASMO: Geology and Structures of the GTS and Grimsel Region*. Nagra Arbeitsbericht NAB 16-27. National Cooperative for the Disposal of Radioactive Waste (Nagra), Wettingen, Switzerland.

Schneeberger, R., Mäder, U.K. and Waber, H.N. 2017. Hydrochemical and isotopic (δ^2H, δ^{18}O, ^{3}H) characterization of fracture water in crystalline rock (Grimsel, Switzerland). *Procedia Earth and Planetary Science*, **17**, 738–741, https://doi.org/10.1016/j.proeps.2016.12.187

Schneeberger, R., Kober, F., Lanyon, G.W., Mäder, U.K., Spillmann, T. and Blechschmidt, I. 2019. *Grimsel Test Site: Revisiting the Site-Specific Geoscientific Knowledge*. Nagra Technical Report NTB 19-01. National Cooperative for the Disposal of Radioactive Waste (Nagra), Wettingen, Switzerland.

Shen, Y., Chapelle, F.H., Strom, E.W. and Benner, R. 2015. Origins and bioavailability of dissolved organic matter in groundwater. *Biogeochemistry*, **122**, 61–78, https://doi.org/10.1007/s10533-014-0029-4

Shettima, A.Y., Karumi, Y., Sodipo, O.A., Usman, H. and Tijjani, M.A. 2013. Gas chromatography-mass spectrometry (GC-MS) analysis of bioactive components of ethyl acetate root extract of *Guiera senegalensis* J.F. Gmel. *Journal of Applied Pharmaceutical Science*, **3**, 146–150, http://dx.doi.org/10.7324/JAPS.2013.30328

Shimizu, S., Akiyama, M., Naganuma, T., Fujioka, M., Nako, M. and Ishijima, Y. 2007. Molecular characterization of microbial communities in deep coal seam groundwater of northern Japan. *Geobiology*, **5**, 423–433, https://doi.org/10.1111/j.1472-4669.2007.00123.x

Simon, A.G., Mills, D.K. and Furton, K.G. 2017. Chemotyping the temporal volatile organic compounds of an invasive fungus to the United States, *Raffaelea lauricola*. *Journal of Chromatography A*, **1487**, 72–76, https://doi.org/10.1016/j.chroma.2017.01.065

Stillings, M., Lunn, R.J., Pytharouli, S., Shipton, Z.K., Kinali, M., Lord, R. and Thompson, S. 2021. Microseismic events cause significant pH drops in groundwater. *Geophysical Research Letters*, **48**, e2020GL089885, https://doi.org/10.1029/2020GL089885

Suzuki, R. and Shimodaira, H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, **22**, 1540–1542, https://doi.org/10.1093/bioinformatics/btl117

US EPA 1996. *Test Methods for Evaluating Solid Waste, Method 3510c, Revision 3 December 1993, Final Update III to the Physical/Chemical Methods*. EPA Publication SW-841. United States Environmental Protection Agency (US EPA), Washington, DC.

US EPA 2017. *Method 3545A (SW-846): Pressurized Fluid Extraction (PFE)*. United States Environmental Protection Agency (US EPA), Washington, DC, https://www.epa.gov/esam/method-3545a-sw-846-pressurized-fluid-extraction-pfe [last accessed 14 July 2020].

US Geological Survey 2006. *CFC Bottle Sampling Method*. United States Geological Survey, Reston, VA, Chlorofluorocarbon Laboratory Website, https://water.usgs.gov/lab/chlorofluorocarbons/sampling/bottles/ [last accessed 29 August 2019].

Zhang, Y., van Dijk, M.A., Liu, M., Zhu, G. and Qin, B. 2009. The contribution of phytoplankton degradation to chromophoric dissolved organic matter (CDOM) in eutrophic shallow lakes: Field and experimental evidence. *Water Research*, **43**, 4685–4697, https://doi.org/10.1016/j.watres.2009.07.024