BTO 2018.072 | July 2018

BTO report

Clustering and prioritisation to design a risk based monitoring program in groundwater sources for drinking water

BTO

Clustering and prioritisation to design a risk based monitoring program in groundwater sources for drinking water

BTO 2018.072 | July 2018

Project number
401876/001/001

Project manager
Stefan Kools

Client
Speerpuntonderzoek - Vitens

Quality Assurance
Annemarie van Wezel

Author(s)
Rosa M.A. Sjerps[a], Andrea M. Brunner[a], Yuki Fujita[a], Bernard Bajema[b], Martin de Jonge[b], Patrick Bauerlein[a], Joost de Munk[a], Merijn Schriks[b]
a=KWR Watercycle Research Institute, b=Vitens

Sent to
This report is distributed to Vitens.
A year after publication it is public.

Year of publishing
2018

More information
Dr. Stefan A.E. Kools
T
E

Keywords

Chemistry, water quality, monitoring

Postbus 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

BTO *Management samenvatting*

Clustering en prioritering voor het ontwerpen van een op risico gebaseerd monitoringprogramma in grondwaterbronnen voor drinkwater

Auteur(s) Rosa M.A. Sjerps[a], Andrea M. Brunner[a], Yuki Fujita[a], Bernard Bajema[b], Martin de Jonge[b], Patrick Bauerlein[a], Joost de Munk[a], Merijn Schriks[b]

a=KWR Watercycle Research Institute, b=Vitens

Het aantal chemische parameters in monitoringprogramma's voor waterbedrijven is in het afgelopen decennium sterk toegenomen. In overeenstemming met de Europese Drinkwaterrichtlijn (EU DWD) richten de waterbedrijven zich op een op maat gesneden en risico gebaseerd monitoringprogramma. In dit project hebben we een dergelijk monitoringprogramma ontwikkeld voor Vitens, dat voornamelijk grondwater als bron gebruikt.

Methode We gebruiken reeds beschikbare gegevens om de Vitens bronnen te clusteren, zowel doelstofgegevens als screeninggegevens op basis van hoge resolutie massaspectrometrie (HRMS). We stellen prioriteiten op basis van (voorlopige) drinkwater richtwaarden of van de ‘threshold of toxicological concern’ (TTC) en suggereren een op risico's gebaseerd monitoringprogramma voor elk cluster van bronnen. Voor de screeningsgegevens prioriteren we de gevonden ‘suspects’ voor verdere bevestiging van de identiteit op basis van semi-kwantitatieve concentraties in combinatie met informatie over in vitro toxiciteit op basis van de ToxCast-database.

Resultaten
Van de 731 gemeten doelstoffen worden er 153 één of meerdere malen gedetecteerd gedurende een periode van vijf jaar. Een fractie (1.398 van de 12.294) van de responsen uit de HRMS screening komt overeen met 3.590 mogelijk te verwachten stoffen die in gebruik zijn of elders eerder aangetroffen (suspects). 108 bronnen zijn vervolgens op basis van zowel de doelstoffen als de suspects ingedeeld in 7 clusters. De relatief schone bronnen met een laag aantal organische chemicaliën en lage concentraties komen voor in gebieden met alle soorten landgebruik. Clusters van bronnen met relatief hogere aantallen chemicaliën en hogere concentraties komen vaak voor daar waar relatief veel oppervlaktewater infiltreert. Voor geperfluoreerde chemicaliën worden 25 van de 691 mogelijk te verwachten stoffen aangetroffen in de bronnen, terwijl 7 suspects ook voorkomen in drinkwater. De identiteit hiervan is nog verder te bevestigen. Voor de doelstoffen vertoont de eenvoudige behandeling de laagste verwijdering, terwijl op sorptie gebaseerde technieken relatief hoge verwijderingsrendementen vertonen. (Voorlopige) richtwaarden ((p)GLV's) zijn beschikbaar voor 45 van de aangetroffen doelchemicaliën, en worden gebruikt voor prioritering voor het bewaken van frequenties. Deze chemicaliën leiden individueel niet tot zorg voor de menselijke gezondheid. We geven prioriteit aan suspects voor verdere identiteitsbevestiging op basis van semikwantitatief voorkomen in geproduceerd water, detectiefrequenties en informatie over toxische potentie op basis van in vitro toxiciteitgegevens. Zodra de identiteiten van deze suspects bevestigd zijn en ze beoordeeld zijn als relevant, kunnen de suspects worden toegevoegd aan doelstofanalyses.

Implementatie De aanpak in dit project geeft een werkbare ‘workflow’ voor risicogebaseerde monitoring voor doelchemicaliën op basis van clusters van grondwaterbronnen. Ook leidt deze aanpak tot inzicht welke nieuwe relevante chemicaliën nadere aandacht behoeven op basis van HRMS screening.

Report

Dit onderzoek is gerapporteerd in BTO 2018.072 | Juli 2018 Clustering and prioritisation to design a risk based monitoring program in groundwater sources for drinking water.

Daarnaast is de tekst aangeboden voor een peer review artikel; Sjerps RMA, Brunner AM, Fujita Y, Bajema B, De Jonge M, Bauerlein P, De Munk J, Schriks M, Van Wezel A (submitted for peer review). Clustering and prioritisation to design a risk based monitoring program in groundwater sources for drinking water.

.

More information
Dr. Stefan A.E. Kools
T +31 30 606 9539
E **Stefan.kools@kwrwater.nl**

KWR
PO Box 1072
3430 BB Nieuwegein
The Netherlands

Summary

The number of chemical parameters included in monitoring programs of water utilities increased in the last decade. In accordance with the European Drinking Water Directive (EU DWD), utilities aim at a tailored risk-based monitoring program. Here, such a risk-based monitoring program is developed for the largest Dutch water utility, mostly using groundwater as a source. We use available data to cluster the different source waters, both target data as non-target/suspect monitoring data based on high resolution mass spectrometry (HRMS). We prioritise targets based on (preliminary) drinking water guideline values or the threshold of toxicological concern (TTC) and suggest a risk based monitoring program for each cluster of source waters. We prioritize the suspects for further identity confirmation based on semi-quantitative occurrence concentrations combined with in vitro toxicity information based on the ToxCast database.

Out of 731 measured target chemicals, 153 are once or multiple times detected over a five year period. A fraction (1,398 out of 12,294) of occurring non-target features matches to 3,590 suspects. 108 source waters are clustered into 7 clusters. Source waters with a low number of organic chemicals which are detected in low concentrations, are located in areas with all land-use types. Clusters of source waters in which higher numbers of chemicals occur are related to high levels of infiltrated surface water. For perfluorinated chemicals, 25 out of 691 suspects match detected features in source waters while 7 suspects are also found in drinking water. For the target chemicals simple treatment shows lowest removal rates, while sorption based techniques show relatively high removal efficiencies. The chemical composition of all drinking waters relates to non-contaminated source waters. (Preliminary) guideline values ((p)GLVs) are available for 45 of the retrieved target chemicals, and are used for prioritisation for monitoring frequencies. These chemicals individually pose no appreciable concern to human health. We prioritize suspects for further identity confirmation based on semi-quantitative occurrence in produced water, detection frequencies and information on toxic potency. Once confirmed and assessed as relevant, the suspects can be added to target monitoring.

This approach provides a feasible workflow for risk based monitoring for target chemicals for clusters of groundwater sources, connected to a feed of new relevant chemicals based on HRMS suspect screening.

Contents

1 Introduction

1.1 Towards risk based monitoring

Worldwide, drinking water regulations prescribe drinking water quality standards for a selection of chemicals. The EU Drinking Water Directive (EU DWD) for example lists standards for 26 chemical parameters. Most drinking water utilities monitor a broad set of parent chemicals and their transformation products, using target, non-target (Hollender et al. 2017) and bioanalytical methods (Leusch et al. 2017). The EU DWD stimulates that drinking water monitoring is performed in a more flexible way, provided that protection of public health is ensured. The aim is to reduce obsolete analyses and concentrate on relevant issues, following the principle of ‘hazard analysis and critical control point’ (HACCP) (Van Wezel et al. 2010) and the water safety plan approach as developed by WHO (Kot et al. 2015).

Compared to surface water, groundwater is less intensively studied and monitored (Loos et al. 2010, Lapworth et al. 2012, Jurado et al. 2012). Groundwater can however be highly influenced by anthropogenic activities related to the land-use (Ter Laak et al. 2012), by infiltrating surface water (Sui et al. 2015), by historical contamination (Eggen et al. 2010) or by activities in the sub-soil (Bonte et al. 2011). The susceptibility of the groundwater aquifers to these pressures depends on soil characteristics and groundwater hydrology (Mendizibal et al. 2012, Van Wezel et al. 2009). Chemical properties, such as persistence and mobility, are reflected in spatio-temporal patterns of chemical occurence in groundwater after emissions. The chemical properties also influence removal efficiencies during drinking water production, depending on the water treatment techniques applied (Van Wezel et al. 2017).

Water utility Vitens services drinking water in a large area in the Netherlands, using groundwater as a major source. The set of organic chemical parameters in their monitoring program tripled the last decade. In accordance with the EU DWD, the water utility aims to prioritize measured chemicals and to develop a tailored risk-based monitoring program. In literature several prioritisation methods for chemicals of emerging concern (CEC) have been developed (Guillén et al., 2012), that make use of target monitoring data (von der Ohe et al., 2011), non-target and suspect screening data (Hollender et al., 2017; Moschet et al., 2014; Sjerps et al., 2016), exposure models (Arnot et al., 2012; Wambaugh et al., 2013) or chemo-informatics (Guha et al., 2016).

The aim here is to develop a risk-based monitoring program for the drinking water sources involved in the service area of the water utility. We use available target and non-target/suspect monitoring data and characteristics of the supply zones. We use clustering techniques to cluster the supply zones based on target and suspect data. We prioritise targets based on (preliminary) drinking water guideline ((p)GLVs) values or threshold of toxicological concern (TTC). Based on this information we suggest a risk based monitoring for each clusters of supply zones. We prioritize the suspects for identity confirmation based on semi-quantitative concentrations combined with in vitro toxicity information.

2 Materials and methods

2.1 Typology drinking water supply zones

The data used originate from 141 source waters, mixed water from one or multiple pumping wells prior to drinking water treatment in the central, eastern and northern parts of the Netherlands. Two drinking water supply zones are mainly fed by river bank filtrate, the other supply zones use groundwater as a source. Per source water the percentage infiltrated surface water is given, expressed in four classes i.e. i) 5-10%, ii) 10-20%, iii) 20-50% and iv) 50-70%. The supply zones are classified following the ABIKOU typology (Stuyfzand 1996, Van Wezel et al. 2009), in which A corresponds to phreatic groundwater in sandy soil, B for (semi-)confined groundwater, I for artificially infiltrated surface water and U for riverbank infiltrated surface water. The land-use in the 25 year infiltration zone is defined as the percentages of urban, agriculture and nature area in the total recharge area.

The water is treated at 96 production stations. The drinking water treatment techniques consist mostly of commonly used drinking water treatment techniques such as flocculation, sand filtration, aeration, water softening, pH adjustment and more occasionally also includes reverse osmosis (RO) and active carbon filtration.

2.2 Analytical chemistry

We use monitoring data generated by Vitens drinking water laboratory. This laboratory works via strictly defined QA/QC criteria, takes part in round robin tests, works via ISO, OECD or NEN standard procedures when available, and is officially accredited via the Dutch Board for Accreditation. Vitens routinely performs monitoring in both the source and produced waters, for 731 target chemicals using several methods (See S.I. RBM.xlsx ‘s.i. targets’). Current monitoring frequency in source water is at least once per year. The frequency depends on the estimated susceptibility of the supply zone and on if the parameter is explicitly mentioned in current legislation. Here we use routine target monitoring data produced in the period 2010 to 2016. This dataset consists of 553,440 entries for source water including 8,954 entries above reporting limits, and 760,339 entries for drinking water including 5,352 entries above reporting limits. For each parameter, the frequency of detection and variability (averages and 90th percentiles) over 2010-2016 of the detected concentrations is deduced averaged over all samples from source waters, and averaged per cluster of source waters and drinking water.

In addition, the source and produced waters of all supply zones are in 2016 once monitored using non-target high-resolution mass-spectrometry (high-pressure liquid chromatography). Vitens is equipped with an AB Sciex Q-TOF (API Triple TOF 5600+), used in positive and negative ionization mode. In total, this dataset consists of 41,267 detected entries in source water and 12,123 detected entries in drinking water. All results are expressed in terms of internal standard equivalent (IS-eqs.), for both positive and negative ionization mode neburon was used as internal standard. A total of 12,294 non-targets features (7,503 using positive ionization mode and 4,791 using negative ionization mode) are matched to NORMAN SusDat (14,632 entries, www.norman-network) and Sjerps et al. (2016) (5,219 entries) for the purpose of suspect screening. The latter consists of industrial chemicals (>100 ton), pharmaceuticals, veterinary pharmaceuticals, pesticides and biocides which are authorized on the European market.

Specific attention is paid to perfluorinated chemicals, for which the Norman PFAS suspect list was used comprising 691 CAS-numbers. Suspect data are filtered according to their accurate mass (tolerance <5ppm) and predicted retention time (tolerance <3 min). In the present study, confidence levels of the retrieved suspects, according to the scheme by Schymanski et al. (2014), are not defined. For each parameter, the frequency of detection and variability (averages and 90th percentiles) of the semi-quantitative concentrations is given.

2.3 Clustering

To cluster the source water samples, average concentrations of each target chemical are calculated over a period of 6 years for each sampling location. Average concentrations are based on detected concentrations above the reporting limit (RL); when all measurements on a sampling location over 2010-2016 are below RL the concentration is expressed as 0.5*RL, based on the lowest RL for the target chemical in the dataset. The following is excluded from the dataset; a) CH_4, DOC, TOC, b) chemicals that are not found above RL in any of the source water samples, c) chemicals that are measured in less than 100 water samples and d) source water samples for which less than 50 chemicals are measured. All chemical concentrations are log-transformed. This results in a subset of 108 source water samples and 152 target chemicals.

The 108 source waters are clustered using k-means clustering. This is a commonly used algorithm of unsupervised learning, and is used to partition a number of observations into k clusters based on their similarity. To relate the clusters of source water to information of a large number of chemicals, we reduce dimensionality of the dataset using principal component analysis (PCA). The chemicals that are detected in only one water sample are excluded. The major axes of variations extracted with PCA are interpreted based on the loading of each chemical. The clusters of source waters are projected on the reduced dimensions of PCA. In addition, the clusters are also projected on a plane of two metrics which represent overall abundance of target chemicals, i.e. total concentrations and number of all detected chemicals. Finally, the clusters of source waters are compared to surface water influence, the proportion of land-use types (urban, agriculture, nature), and the ABIKOU class. For the sake of presentation, the clusters are numbered based on their median values of total concentration of all detected chemicals.

Identically, drinking water samples are also clustered based on target chemicals. The above mentioned exclusions result in a subset of 101 drinking water samples and 112 target chemicals. Chemicals that are detected in only one water sample are excluded, leaving 72 chemicals. K-means clusters are related to treatment class applied to each drinking water (Table 1).

Using the PCA loadings of detected target chemicals in source water, the PCA scores of 101 drinking water samples are calculated and plotted on the PCA plane based on source water. The PCA scores of drinking water are derived by multiplying the concentrations of target chemicals in drinking water with the PCA loadings computed from target chemicals in source water. Known pairs of source water and produced drinking water are connected by arrows. In this way, the chemical composition of drinking water can be projected on the same 2D plane as source water, enabling a visualization of change in water quality due to treatment.

Source water samples are also clustered based on suspect chemicals. After the same exclusion procedure as target chemicals, source water samples and 1,297 suspect chemicals are used for k-means clustering. Prior to the analysis, suspect chemical concentrations are log-transformed after adding 0.0001 µg/L IS-eq. Since the number of suspects is too large compared to the number of water samples to conduct PCA, we reduce the number of suspects from 1,297 to 162 by selecting only those that are detected in more than 5 water samples and with 90th percentiles greater than 0.01 µg/L IS-eq (see Figure S.I.1).

All statistical analyses are conducted using R version 3.4.1.

2.4 Analysis of treatment efficiencies

Removal efficiencies are derived for all detected target chemicals in source water for locations with comparable combination of treatment techniques. For each drinking water production location and per target chemical, individual measurement of the concentration in the (mixed) source water is compared to the corresponding individual measurement of the concentration in the produced drinking water. The calculated removal efficiencies are expressed per group of production locations with similar treatment techniques (Table 1) and over all production locations. For parameters for which concentrations in drinking water are <RL, RL is assumed as a realistic worst case approach. Removal efficiencies are calculated as $(C_{source} - C_{drinking\ water})/(C_{source})$.

Table 1. Treatment technology classes of the drinking water production locations.

Treatment class	Rapid sand filtration / marble filtration	Active carbon filtration	Nanofiltration or reversed osmisis	# production locations
Simple	x	-	-	81
Sorption	x	x	-	4
Size excusion	x	-	x	10
Sorption combined with size exclusion	x	x	x	3

2.5 Prioritisation and risk based monitoring for target chemicals

When (preliminary) drinking water guideline values ((p)GLVs) are available for target chemicals present in source waters or produced drinking water, these are used for further prioritisation (eg. Baken et al. submitted). Chemicals are prioritized for all supply zones and per cluster by comparing averages and 90th percentiles of the concentration in source water and produced drinking water to the (p)GLVs. The ratio of both is expressed as the Benchmark Quotient (BQ, Schriks et al. 2010). For those target chemicals for which no (p)GLVs are available, the concentrations in produced drinking and source water are compared to the TTC (threshold of toxicological concern) value (Mons et al., 2013).

We suggest that all target chemicals that are not detected in any source or produced water, can be monitored in a lower frequency, in accordance to the monitoring obligations related to the EU Water Framework Directive. Higher frequencies are recommended for all chemicals that are found in produced or source water, according to Table 2. This risk-based monitoring program for target chemicals is defined per cluster of source waters based on the criteria for monitoring frequency.

Table 2. Criteria for frequency of monitoring of target chemicals in source and drinking water.

Criteria for target chemicals	**Advice for frequency in monitoring program**
Non-detect	Low frequency
Detected only in source water not in produced water, BQ <0,001 or <0,01 µg/L for targets without (p)GLV	Moderate low frequency
Detected in produced drinking water, BQ<0,001 or <0,01 µg/L for targets without (p)GLV	Moderate high frequency
Other	High frequency

2.6 Prioritisation for identity confirmation for suspect chemicals

For both target and suspect chemicals, octanol water partition coefficient (log Kow) and half-life (DT50) values are gathered via EPI Suite (US EPA, 2012). When available, experimental data are preferred over modelled data. DT50 values are predicted according to Biowin 3, which is built on measured biodegradability data of over 200 substances for which molecular fragments are described. Likely biodegradation half-lives are expressed by a score system, i.e. 5 reflects hours, 4 reflects days, 3 reflects weeks, 2 reflects months and 1 reflects years (Aronson et al. 2006). For further analysis of the suspects in relation to log Kow and DT50, all features that match to more than 5 suspects are neglected for further analyses, to reduce uncertainty.

For further analysis of the suspects in relation to their toxicity, features that match a similar suspect from the different suspect lists are reduced to one entry. Minimum and 5th percentile AC50 values, i.e. the concentration at which 50% of the maximum response is achieved per chemical per in vitro bioassay, are gathered from EPA's ToxCast database (e.g. Richard et al. 2016, US EPA, 2015). ToxCast chemical codes are linked to CAS numbers of the suspects retrieved. AC50 values are extracted for all in vitro assays in which a chemical is tested. For more details we refer to Brunner et al. (submitted). The features are prioritized for further confirmation based on the ratio of average IS-eq occurrence in all produced waters divided by the minimum AC50 per feature.

3 Results and discussion

3.1 Clustering of source waters based on targets

Out of 731 measured target chemicals, 153 chemicals are once or multiple times detected.

PCA axes 1 and 2 of target chemicals in source water explain respectively 14,8% and 9,4% of the total variance. Axis 1 is associated with negative loading of almost all chemicals and therefore reflects cleanness of water (Figure S.1.2.a). This axis is highly and negatively correlated with the number of detected target chemicals (spearman correlation coefficient $\rho = -0.64$, $p < 0.001$) and the total concentrations of the target chemicals ($\rho = -0.46$, $p < 0.001$). Source water which is influenced with a large amount of surface water scores low on this axis (Figure S.I.2.a). PCA axis 2 reflects the type of chemicals present in the sample, since most of the pesticides, pharmaceuticals or artificial sweeteners are positively related to this axis while industrial chemicals are negatively related (Figure S.I.2.b). Accordingly, the scores of samples on this axis are positively correlated with the proportion of agricultural land-use ($\rho = 0.44$, $p < 0.001$) and negatively correlated with the proportion of urban land-use ($\rho = -0.63$, $p < 0.001$).

The clustering of the source waters based on target chemicals is depicted in Figure 1a and Figure 1b (see S.I.RBM.xlx ‘s.i. sources’) for clustering of the individual source waters and their properties). A k-value of 7 is chosen because the variance explained by the clusters starts to plateau at k-values between 7 and 10. Cluster 7, which are the relatively non-vulnerable source waters with low concentrations and low number of target chemicals, occurs in all land-use types. Source waters consisting solely of the land-use nature are clustered into cluster 7. Source waters in cluster 3 and 4, in which higher number of target chemicals are found, consist of more than 50% of infiltrated surface water. Two wells influenced by point source contamination with chlorinated hydrocarbons are separately clustered in cluster 1. See Figure S.I.3. for more information on clustering of source waters related to the supply zone typology in terms of land-use and influence of surface water infiltration.

3.2 Clustering of source waters based on suspects

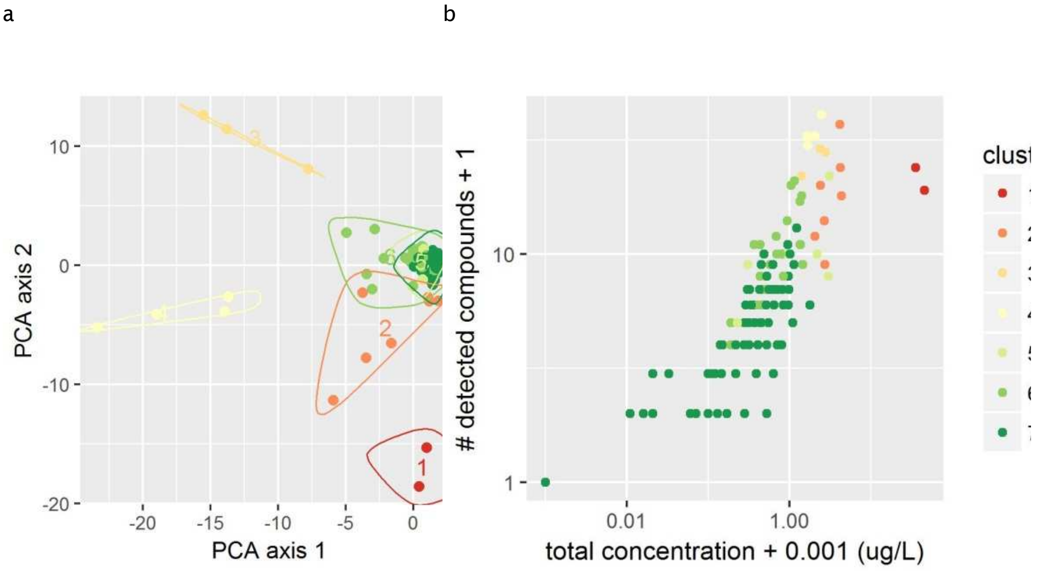
In all 141 individual source waters, 1,398 features are retrieved that match to 3,590 suspects as described. Detected suspects do not show a different pattern in hydrophobicity of toxicity compared to non-detected suspects (Fig. S.I.4.). Features can match to a maximum of 36 different suspects, on average features match to 3 different suspects both in the positive and negative ionisation mode (Fig. S.I.5). The majority of the suspects retrieved will therefore be false positives. Using smaller suspect lists will lead to fewer hits and fewer false positives, but potentially also to false negatives. Similar, to clustering based on the detected target chemicals, 7 clusters of source waters were distinguished based on the detected suspects (Figure 1c & 1d, S.I.RBM.xlsx).

PCA axis 1 and 2 of suspect chemicals in source water explain respectively 19.8% and 7.4% of the total variance (Figure 1c & d). For source waters in cluster 7, again the relatively non-vulnerable source waters, all land-use types are present in their recharge areas; however recharge areas with a high proportion of agricultural area are less frequently present. A high number of suspect chemicals are found in source waters from cluster 1, 3 and 4, influenced by more than 50% infiltrated surface water. See Figure S.I.3. for more information on clustering of source waters related to typology in terms of land-use, influence of surface water infiltration and structure of the subsoil.

A comparison of clustering based on target and suspect chemicals (Table 3) shows that approximately half of the source waters (56 out of 108, grouped as cluster 7 for both) can be considered as relatively non-vulnerable to anthropogenic influences in terms of both target chemical composition and non-target chemical composition. Seven of the source waters, i.e. cluster 3 and 4 for the targets and cluster 1,3, and 4 of the suspects, are similar with relatively high levels of surface water infiltration. There is an large overlap between cluster 6 based on targets and cluster 6 based on suspects, which consists of source waters with a high percentage of agricultural land-use.

However it is also clear that suspect screening gives complementary information to the target analyses (Sjerps et al. 2016), as many other source waters clustered differently based on either target or suspect data. An example are 5 source waters from cluster 7 based on the suspects, consisting of relative clean waters, that occur in cluster 1 and 2 according to the targets, consisting of relatively contaminated waters. On the other hand, 5 source waters from cluster 1 and 2 based on the suspects, consisting of relatively contaminated water also cluster in clusters 6 and 7 based on the targets, consisting of relatively clean waters. An explanation for these differences is chemicals that are not well ionized or that are very volatile cannot easily be detected via liquid chromatography high resolution mass spectrometry (LC-HRMS) used for suspect screening.

Target Chemicals



Suspect chemicals

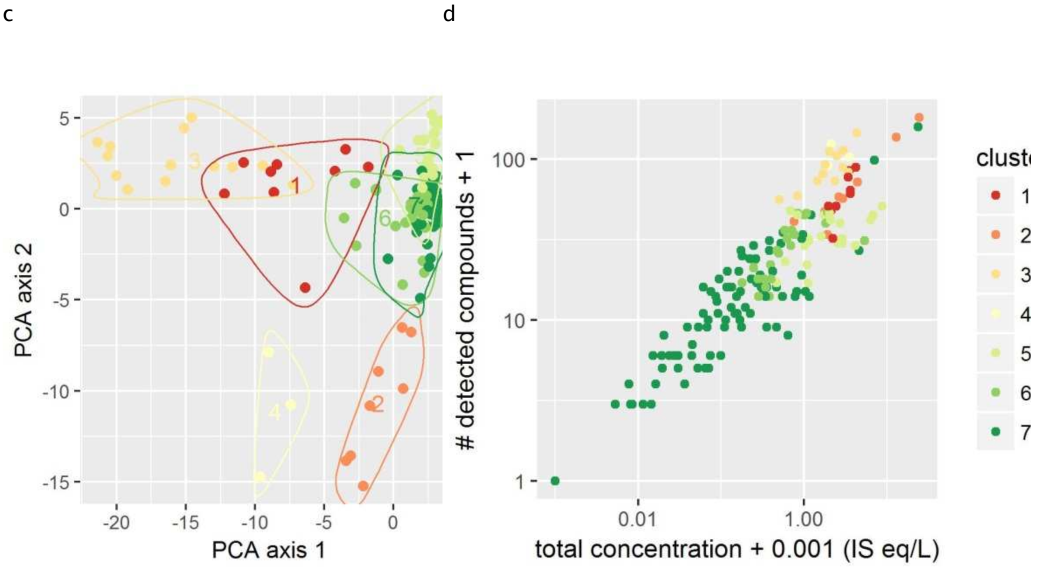


Figure 1: Clustering of source water target data (a) and suspect data (c), plotted on PCA axis 1 and 2, and plotted according to total concentration and number of detected chemicals per sample for target data in µg/L (b) and suspect data in IS eq/L (d).

Table 3: Clustering of 108 out of 141 source waters based on target chemicals and suspect chemicals compared (see also SI RBM.xlsx ´S.I. sources´)

Target based cluster →	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Sum**
Suspect based cluster ↓								
1	0	0	2	0	0	1	0	3
2	0	3	0	1	0	2	2	8
3	0	0	1	0	0	0	0	1
4	0	0	0	3	0	0	0	3
5	0	1	0	0	3	1	7	12
6	0	0	0	0	1	9	2	12
7	2	3	0	0	3	5	56	69
Sum	2	7	3	4	7	18	67	108

3.3 Perfluorinated chemicals

For the perfluorinated chemicals, 25 suspects from the Norman PFAS suspect list match features in the source waters. Depending on the exact suspect 1 to 33 different supply zones for source water contain these suspects, while 7 suspects are also retrieved in drinking water, in 1 to 14 different production stations (Table S.I.1). Merely four of these 25 retrieved suspect perfluorinated chemicals are REACH registered. For 17 chemicals the registration status is “pre-registered”. For these chemicals information on which companies are actually producing/using them cannot be retrieved. Furthermore, only for a few chemicals it is known what they are actually used for. They are mainly employed as surfactants. A total of 14 of these chemicals could not be found as mentioned in scientific literature. However, there are two papers dealing with the global emission of several C4-C14 PFCA’s (Wang et al. 2014ab). In Korea mean concentrations of PF’s in WWTP effluent and sludge are between 1 ng/L – 800 ng/L and 1 – 100 ng/g (Kwon et al. 2017). At least two chemicals from the list are found in WWTP’s in Korea: 5 ng/L (355-46-4) and 80 ng/L (335-67-1) (Kwon et al. 2017).

3.4 Analysis of treatment efficiencies

When samples of produced drinking water are plotted on the PCA planes derived from target chemicals in source water (Figure 1a), they coincide with cluster 7 of the non-vulnerable source waters (Figure 2). Both simple and sorption techniques, combined with mixing of individual source waters, have a positive effect on the composition of the water quality.

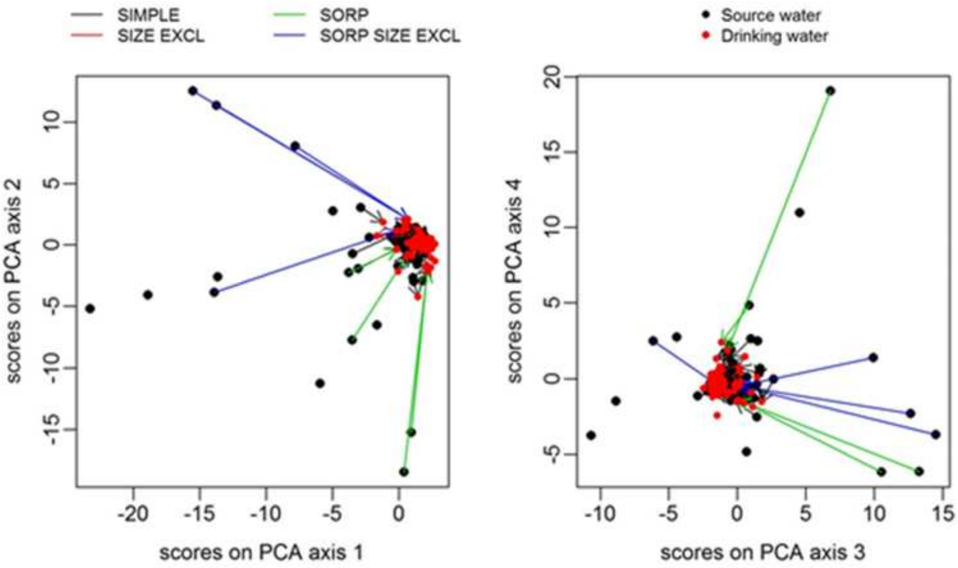


Figure 2. PCA scores for source water (black) and drinking water samples (red) plotted on the PCA axes as derived in Figure 1a. The source water is connected to the complementary drinking water by lines.

The mean removal efficiencies for simple, sorption and size exclusion treatment techniques differ significantly (ANOVA, $p<0.01$, Figure 3a). Variability in removal efficiency within locations with the same treatment techniques does occur. Drinking water treatment based on only simple treatment techniques shows as expected the lowest removal rates, while sorption based techniques -granulated activated and powder activated charcoal- show relatively high removal efficiencies. Techniques for size exclusion include reverse osmosis and nanofiltration and generally treat only half of the drinking water volume at the production locations of the water utility and followed by mixing with differently treated water. The removal rates presented in Figure 3a are based on concentrations in mixed drinking water which explains the relatively low removal efficiencies. Removal efficiencies for target chemicals treated with sorption techniques, i.e. active carbon filtration, show as expected (Westerhoff et al. 2005) a significant correlation with hydrophobicity ($p<0.01$, Figure 3b), however the explained variance is low ($R2=0.02$). Several target chemicals, at few points in time and at few production locations, are introduced or show an increase in concentration during drinking water treatment as a result of transformation processes. This holds for 65 chemicals and for 69 production locations.

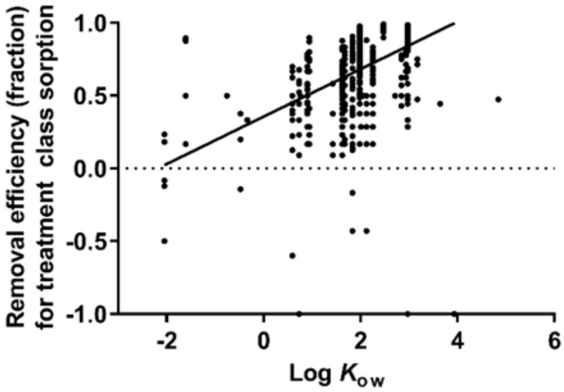


Figure 3. (a.) Distribution of removal efficiencies, including removal efficiencies based on <RL in drinking water, for target chemicals per treatment type. Box extends from 25th to 75th percentiles and whiskers extent from 1 to 99th percentiles, size exclusion is applied on only half of the produced drinking water volume. (b.) Relation between removal efficiencies for individual target chemicals for production stations where sorptive techniques are included and hydrophobicity ($p<0.01$, $R^2=0.02$).

3.5 Prioritisation and risk based monitoring for target chemicals

For the prioritisation of target chemicals (provisional) drinking water guideline values ((p)GLVs) are used, which are available for 45 of the 153 target chemicals found in source and drinking water. For all these target chemicals, concentrations in drinking water are below the benchmark quotient of 0.1 (Figure 4). So, these individual target chemicals pose individually no appreciable concern to human health, which is in line with earlier conclusions (Schriks et al. 2010, Baken et al. submitted, Bruce et al. 2010, De Jongh et al. 2012, Houtman et al. 2014).

In drinking water, 19 chemicals with a pGLV and 22 chemicals without an pGLV have a BQ>0.001 based on the 90th percentile concentration. According to Table 2, these chemicals are advised to be most frequently monitored in drinking and source water. For source water, 32 chemicals with an available pGLV and 81 chemicals without an pGLV have a BQ>0.001 based on the 90th percentile concentration. Again, these chemicals are advised to be most frequently monitored in drinking and source water, and when possible to derive a pGLV if this is absent. For each cluster of source waters, according to the established criteria for frequency of monitoring of target chemicals in source and drinking water (Table 2), a suggestion for a risk based monitoring program for the target chemicals is given (Table SI RBM.xlsx sheet 's.i. targets').

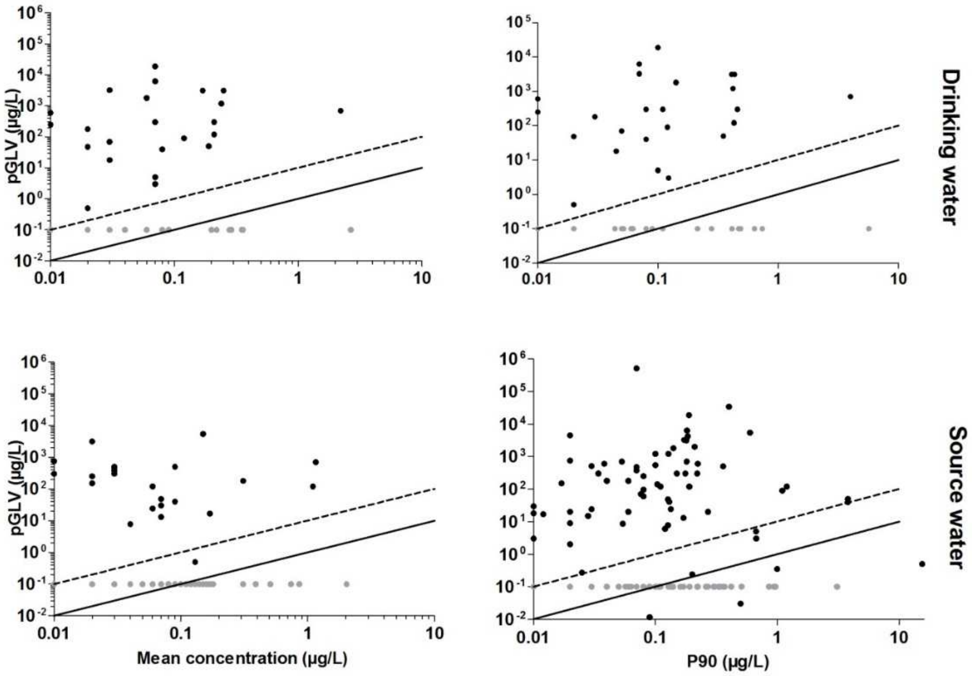


FIGURE 4. PROVISIONAL DRINKING WATER GUIDELINE VALUES (BAKEN ET AL. SUBMITTED) COMPARED TO MEAN AND 90TH PERCENTILE CONCENTRATIONS FOUND IN DRINKING WATER AND SOURCE WATER. CONCENTRATIONS OF CHEMICALS WITHOUT PGLVS ARE COMPARED TO OF THE TTC VALUE OF 0.1 µG/L (IN GREY). BLACK LINE REPRESENTS A BENCHMARK QUOTIENT OF 1, WHILE DOTTED LINE REPRESENTS AN BENCHMARK QUOTIENT OF 0.1

3.6 Prioritisation for identity confirmation for suspect chemicals

As features can match multiple suspects, further effort is needed to confirm identity based on e.g. isotopic patterns and MS2 fragmentation data (Schymanski et al. 2014) and ultimately by obtaining a reference standard and match the retention time and spectra. In view of the efforts demanded, automation of structural identification based on MS2 data, cross-laboratory exchange of information and open science will be needed to achieve this (Schymanski and Williams, 2017). Structured, semi-automated workflows are being developed for prioritisation and confirmation (Pochodylo and Helbling 2017, Kaserson et al. 2017, Gros et al. 2017, Hollender et al. 2017).

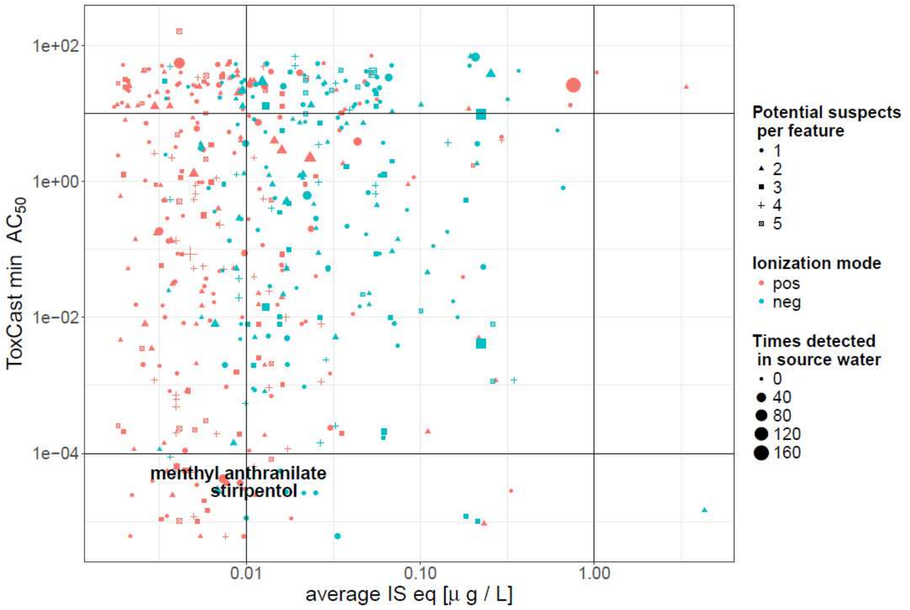
Here we prioritize suspects for which it is warranted to further confirm identity. Once confirmed and assessed as relevant the suspects can be added to the target monitoring, as the semi-quantitative expression of concentrations in IS-eq. brings along large uncertainties of multiple orders of magnitude related to the concentrations as expressed based on reference standards (Sjerps et al. 2016). After a period of more intensive monitoring to collect a sufficient body of data, again prioritisation and risk based monitoring can be performed as described.

Of the 3,590 retrieved suspects 1,017 have a type of use classification (Sjerps et al. 2016), and for 2,398 and 2,819 of the suspects information is available on respectively log Kow and DT50 according to EPI Suite (US EPA, 2012). For 2,400 of the retrieved suspects, AC50 data are available in the EPA ToxCast database.

Average concentrations and frequencies of detection in relation to log Kow and DT50 show no clear pattern that more hydrophobic and degradable suspects are better removed (Fig. S.I.7). Such a pattern would be expected (Reemtsma et al. 2016), but the relation is probably disturbed by false positives occurring in the dataset.

Data on average I.S.-eq. and AC50 values per feature are given in Figure 5, for source and produced drinking water. The number and concentrations of suspects are as expected higher in source water as compared to produced drinking water. Many suspects retrieved in the source waters are not found in finished drinking water. Only a limited number of suspects is found in finished drinking water but not in the source water, potentially transformation products formed during drinking water production (Bader et al. 2017). This will be further detailed in a separate study.

The suspects are prioritized for further confirmation based on the ratio of average IS-eq occurrence in produced water divided by the minimum AC50 times the detected frequency (Table 4 and SI RBM.xlsx sheet ‘s.i. suspects’). For a feature of which the suspect is to be confirmed, all possible suspects for that feature are to be considered.



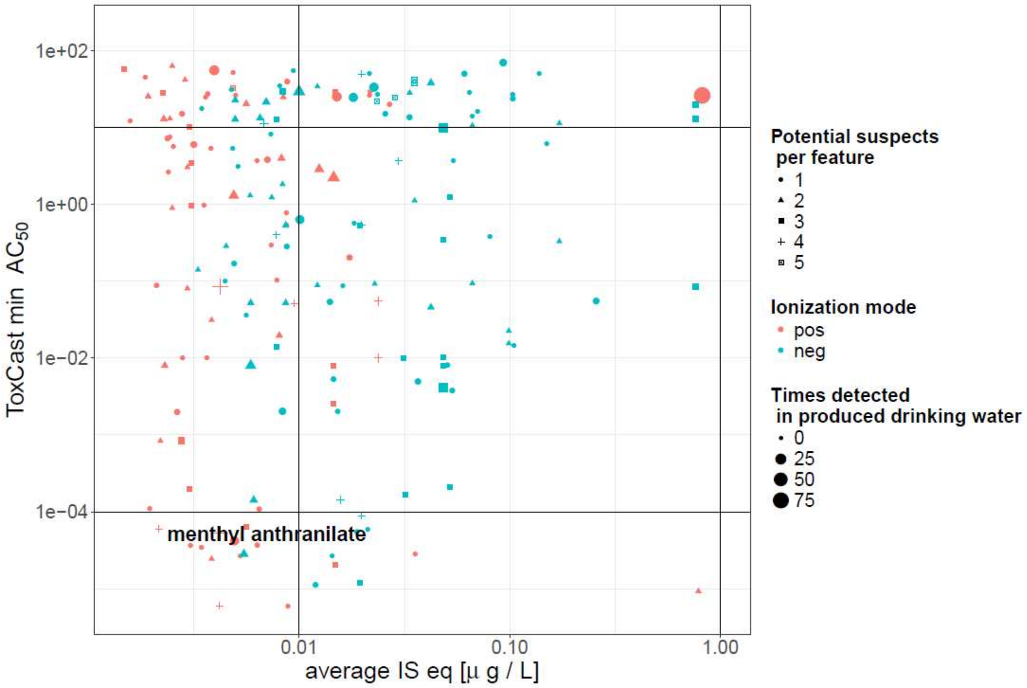


FIGURE 5. AVERAGE SUSPECT CONCENTRATION VERSUS IN VITRO TOXICITY AS BASED ON MINIMUM AC50 FOR SUSPECTS IN SOURCE (A) AND PRODUCED (B) WATER.

Table 4. Top 20 prioritized suspects for further confirmation of identity.

ionisation mode	Feature (m/z / RT)	CAS	Suspect	Frequency of detection	Ratio average concentration in IS-eq / minimum AC50
+	114.0913 / 12.54	2687-91-4	1-ethylpyrrolidin-2-one	1.1%	84794
+	286.0724 / 16.65	4291-63-8	cladribine	18.1%	116
+	276.1967 / 29.87	134-09-8	menthyl anthranilate	31.9%	80
+	180.1255 / 15.48	680-31-9	hexamethylphosphoric triamide; hexamethylphosphoramide	2.1%	170
+	301.2155 / 27.50	72-63-9	metandienone	1.1%	692
+	221.1538 / 29.00	719-22-2	2,6-Di-tert-butylquinone	3.2%	62
+	297.1856 / 28.40	57-63-6	ethinylestradiol	1.1%	157
+	357.2408 / 30.78	979-32-8	b-Estradiol-17-valerat	1.1%	1477
+	102.0912 / 12.88	109-02-4	4-methylmorpholine	1.1%	734
+	343.1547 / 22.42	27138-31-4	oxydipropyl dibenzoate	1.1%	196
-	190.1252 / 22.16	15299-99-7	napropamide	2.1%	4376
-	233.1191 / 26.53	49763-96-4	stiripentol	8.5%	194
-	321.1688 / 28.30	26538-44-3	Zeranol (Alfa Zearalanol, a-ZAL)	3.2%	1072
-	251.1291 / 23.62	2386-87-0	7-oxabicyclo[4.1.0]hept-3-ylmethyl 7-oxabicyclo[4.1.0]heptane-3-carboxylate	1.1%	541
-	265.1592 / 25.63	126-71-6	triisobutyl phosphate[1]	1.1%	249
-	219.9843 / 14.47	88-51-7	4-amino-6-chlorotoluene-3-sulphonic acid	2.1%	342
-	219.1030 / 21.94	77-83-8	ethyl 2,3-epoxy-3-phenylbutyrate	6.4%	43
-	137.0256 / 14.34	99-06-9	3-hydroxybenzoic acid	1.1%	223
-	232.9797 / 18.48	120-36-5	dichlorprop	1.1%	353

[1] *triisobutyl phosphate has a similar mass as tributylphosphate detected in concentrations up to 0.2 µg/L*

4 Conclusion

- We propose a feasible workflow to design risk based monitoring for drinking water utilities. The monitoring program is specified for target chemicals for clusters of groundwater supply zones, connected to a feed of new relevant chemicals based on LC-HRMS suspect screening.
- Out of 731 measured target chemicals, 153 chemicals are once or multiple times detected in all sources and produced drinking waters over a five year period.
- 1,398 out of 12,294 occurring non-target HRMS features match to 3,590 suspects. Detected suspects do not show a different pattern in hydrophobicity of toxicity compared to non-detected suspects. Many suspects retrieved in the source waters are not found in finished drinking water, while only a limited number of suspects is found in finished drinking water but not in the source water. We prioritized suspects for which the identity is to be further confirmed based on the ratio of occurrence in produced water and potency. Once confirmed and assessed as relevant the suspects can be added to the target monitoring.
- 108 source waters are clustered based on target and suspect information in 7 clusters. Approximately half of the source waters can be considered as relatively non-vulnerable to anthropogenic influences. Clusters of source waters where higher number of chemicals are detected relate to high levels of infiltrated surface water. The chemical composition all drinking waters clusters similar to the non-contaminated sources.
- For perfluorinated chemicals, 25 out of 691 suspects match features in source waters. 7 suspects are also retrieved in drinking water. Limited information is available for the 25 retrieved suspect perfluorinated chemicals both in the EU REACH registration and in scientific literature.
- Produced drinking water clusters with the non-vulnerable source waters. Both simple and sorption techniques, combined with mixing of individual source waters, have a positive effect on the composition of the water quality. Mean removal efficiencies for simple, sorption and size exclusion drinking water treatment technologies differ significantly. Treatment based on only simple treatment shows lowest removal rates, while sorption based techniques show relatively high removal efficiencies.
- For prioritisation of target chemicals, (p)GLVs are available for 45 of the153 retrieved chemicals. These chemicals pose individually no appreciable concern to human health.
- Per cluster of source waters, according to proposed risk-based criteria for frequency of monitoring of target chemicals in source and drinking water, a suggestion for a risk-based monitoring program for target chemicals is given.

5 References

Arnot JA, Brown TN, Wania F, Breivik K, McLachlan MS. 2012. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. Environ. Health Persp. 120:1565-1570.

Aronson D, Boethling R, Howard P, Stiteler W. 2006. Estimating biodegradation half-lives for use in chemical screening. Chemosphere 63:1953-1960.

Bader T, Schulz W, Kümmerer K, Winzenbacher R. 2017. LC-HRMS data processing strategy for reliable sample comparison exemplified by the assessment of water treatment processes. Anal. Chem. 89:13219-13226.

Baken KA, Sjerps RMA, Schriks M, Van Wezel AP (in press) Toxicological relevance and Threshold of Toxicological Concern (TTC) for drinking water relevant contaminants of emerging concern. Environ. Int.

Bonte M, Stuyfzand PJ, Hulsmann A, Van Beelen P. 2011. Underground thermal energy storage: Environmental risks and policy developments in the Netherlands and European Union. Ecol. Soc. 16(1)

Bruce GM, Pleus RC, Snyder SA. 2010. Toxicological relevance of pharmaceuticals in drinking water. Environ. Sci. Technol. 44:5619-5626.

Brunner AM, Dingemans MM, Baken K, Van Wezel AP. Prioritizing anthropogenic chemicals in drinking water and sources through combined use of mass spectrometry and Toxcast toxicity data. J. Haz. Mat.

De Jongh CM, Kooij PJF, De Voogt P, Ter Laak TL. 2012. Screening and human health risk assessment of pharmaceuticals and their transformation products in Dutch surface waters and drinking water. Sci. Tot. Environ. 427-428:70-77.

Eggen T, Moeder M, Arukwe A. 2010. Municipal landfill leachates: A significant source for new and emerging pollutants. Sci. Tot. Environ. 408:5147-5157.

Gros M, Blum KM, Jernstedt H, Renman G, Rodríguez-Mozaz S, Haglund P, Andersson PL, Wiberg K, Ahrens L. 2017. Screening and prioritisation of micropollutants in wastewaters from on-site sewage treatment facilities. J. Haz. Mat. 328:37-45.

Guha N, Guyton KZ, Loomis D, Barupal DK. 2016. Prioritizing chemicals for risk assessment using chemoinformatics: Examples from the IARC monographs on pesticides. Environ. Health Persp. 124:1823-1829.

Guillén D, Ginebreda A, Farré M, Darbra RM, Petrovic M, Gros M, Barceló D. 2012. Prioritisation of chemicals in the aquatic environment based on risk assessment: Analytical, modeling and regulatory perspective. Sci. Total Environ. 440: 236-252.

Hollender J, Schymanski EL, Singer HP, Ferguson PL. 2017. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ. Sci. Technol. 51:11505-11512.

Houtman CJ, Kroesbergen J, Lekkerkerker-Teunissen K, Van der Hoek JP. 2014. Human health risk assessment of the mixture of pharmaceuticals in Dutch drinking water and its sources based on frequent monitoring data. Sci. Tot. Environ. 496:54-62.

Jurado A, Vàzquez-Suñé E, Carrera J, López de Alda M, Pujades E, Barceló D. 2012. Emerging organic contaminants in groundwater in Spain: A review of sources, recent occurrence and fate in a European context. Sci. Tot. Environ. 440:82-94.

Kaserzon SL, Heffernan AL, Thompson K, Mueller JF, Gomez Ramos MJ. 2017. Rapid screening and identification of chemical hazards in surface and drinking water using high resolution mass spectrometry and a case-control filter. Chemosphere 182:656-664.

Kot M, Castleden H, Gagnon GA.2015. The human dimension of water safety plans: A critical review of literature and information gaps. Environ. Rev. 23:24-29.

Kwon HO, Kim HY, Park YM, Seok KS, Oh JE, Choi SD. 2017. Updated national emission of perfluoroalkyl substances (PFASs) from wastewater treatment plants in South Korea. Environ. Poll. 220 A:298-306.

Lapworth DJ, Baran N, Stuart ME, Ward RS. 2012. Emerging organic contaminants in groundwater: A review of sources, fate and occurrence. Environ. Poll. 163:287-303.

Leusch FDL, Neale PA, Hebert A, Scheurer M, Schriks MCM. 2017. Analysis of the sensitivity of in vitro bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. Environ. Int. 99:120-130.

Loos R, Locoro G, Comero S, Contini S, Schwesig D, Werres F, Balsaa P, Gans O, Weiss S, Blaha L, Bolchi M, Gawlik BM. 2010. Pan-European survey on the occurrence of selected polar organic persistent pollutants in ground water. Wat. Res. 44: 4115-4126.

Mendizabal I, Baggelaar PK, Stuyfzand PJ. 2012. Hydrochemical trends for public supply well fields in The Netherlands (1898-2008), natural backgrounds and upscaling to groundwater bodies. J. Hydrol. 450-451:279-292.

Mons MN, Heringa MB, Van Genderen J, Puijker LM, Brand W, Van Leeuwen CJ, Stoks P, Van der Hoek JP, Van der Kooij D. 2013. Use of the Threshold of Toxicological Concern (TTC) approach for deriving target values for drinking water contaminants. Wat. Res. 47:1666-1678.

Pochodylo AL, Helbling DE. 2017. Prioritisation of suspect hits in a sensitive suspect screening workflow for comprehensive micropollutant characterization in environmental samples. Environ. Sci.: Wat. Res. Technol. 3:54-65.

Reemtsma T, Berger U, Arp HPH, Gallard H, Knepper TP, Neumann M, Quintana JB, De Voogt P. 2016. Mind the gap: Persistent and mobile organic compounds - Water contaminants that slip through. Environ. Sci. Technol. 50:10308-10315.

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF, Knudsen TB, Kancherla J, Mansouri K, Patlewicz G, Williams AJ, Little SB, Crofton KM, Thomas RS. 2016. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem. Res. Toxicol. 29:1225-1251.

Schriks, M.; Heringa, M.B.; Van der Kooi, M.; De Voogt, P.; Van Wezel, A.P. (2010) Toxicological relevance of emerging contaminants for drinking water quality. Water Res. 44:461-476.

Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J. 2014. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Environ. Sci. Technol. 48:2097-2098.

Schymanski EL, Williams AJ. 2017Open science for identifying "Known unknown" chemicals. Environ. Sci. Techn. 51:5357-5359.

Sjerps RMA, Vughs D, Van Leerdam JA, Ter Laak TL, Van Wezel AP. 2016. Data-driven prioritisation of chemicals for various water types using suspect screening LC-HRMS. Wat. Res. 93:254-264.

Stuyfzand PJ. 1996. Salinization of drinking water in the Netherlands: anamnesis, diagnosis and remediation. SGU Rapporter och Meddelander 87, Proc. 14th SWIM, 17-21 June 1996, Malmö, Geol. Survey Sweden, Uppsala, pp. 168-177.

Sui Q, Cao X, Lu S, Zhao W, Qiu Z, Yu G. 2015. Occurrence, sources and fate of pharmaceuticals and personal care products in the groundwater: A review. Emerg. Contam. 1:14-24.

Ter Laak TL, Puijker LM, Van Leerdam JA, Raat KJ, Kolkman A, De Voogt P, Van Wezel AP. 2012. Broad target chemical screening approach used as tool for rapid assessment of groundwater quality. Sci. Tot. Environ. 427-428:308-313.

U.S. EPA 2012. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA. 2010-2012. EPI Suite.

U.S. EPA. 2015. Toxicity ForeCaster (ToxCast™) Data. https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data. Accessed 16 November 2017.

Von der Ohe PC, Dulio V, Slobodnik J, De Deckere E, Kühne R, Ebert RU, Ginebreda A, De Cooman W, Schüürmann G, Brack W. 2011. A new risk assessment approach for the prioritisation of 500 classical and emerging organic microcontaminants as potential river basin specific pollutants under the European Water Framework Directive. Sci. Tot. Environ. 409:2064-2077.

Van Wezel AP, Puijker L, Vink C, Versteegh A, De Voogt P. 2009. Odour and flavour thresholds of gasoline additives (MTBE, ETBE and TAME) and their occurrence in Dutch drinking water collection areas. Chemosphere 76:672-676.

Van Wezel AP, Mons M, Van Delft W. 2010. New methods to monitor emerging chemicals in the drinking water production chain. J. Environ. Monit. 12:80-89.

Van Wezel AP, Ter Laak TL, Fischer A, Bäuerlein PS, Munthe J, Posthuma L. 2017. Operationalising solutions-focused risk assessment; mitigation options for chemicals of emerging concern in surface waters. RSC Environ. Sci. Water Res. Tech. 3:403 – 414.

Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, Joliet O, Frame A, Rabinowitz J, Knudsen TB, Judson RS, Egeghy P, Vallero D, Cohen Hubal EA. 2013. High-throughput models for exposure-based chemical prioritisation in the ExpoCast project. Environ. Sci. Technol. 47:8479-8488.

Wang Z, Cousins IT, Scheringer M, Buck RC, Hungerbühler K. 2014. Global emission inventories for C4-C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030. Part I: Production and emissions from quantifiable sources. Environ. Int. 70:62-75.

Wang Z, Cousins IT, Scheringer M, Buck RC, Hungerbühler K. 2014. Global emission inventories for C4-C14 perfluoroalkyl carboxylic acid (PFCA) homologues from 1951 to 2030. Part II: The remaining pieces of the puzzle. Environ. Int. 69:166-176.

Westerhoff P, Yoon Y, Snyder S, Wert E. 2005. Fate of endocrine-disruptor, pharmaceutical, and personal care product chemicals during simulated drinking water treatment processes. Environ. Sci. Technol. 39:6649-6663.

6 Attachment(s)

Attachment I - Supporting Information & Figures

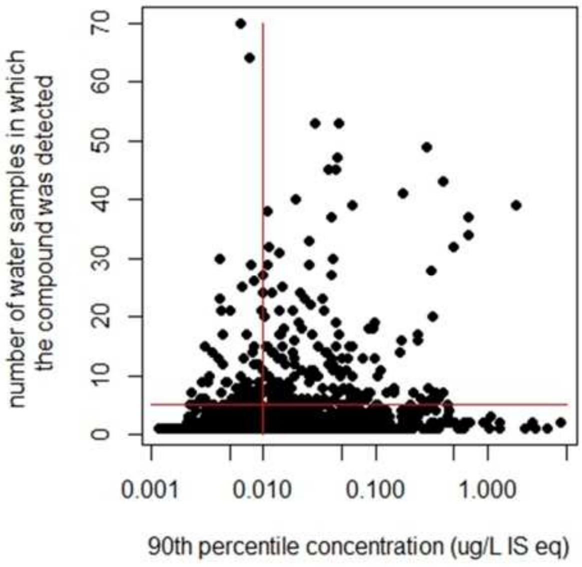


Figure S.I.1. 90th percentile of **suspect** concentration and number of water samples in which the suspect chemical was detected in source water. For PCA analysis for suspects in source water, 162 suspects were used which were detected in more than 5 water samples and have a 90th percentiles >0.01 ug/L IS eq.

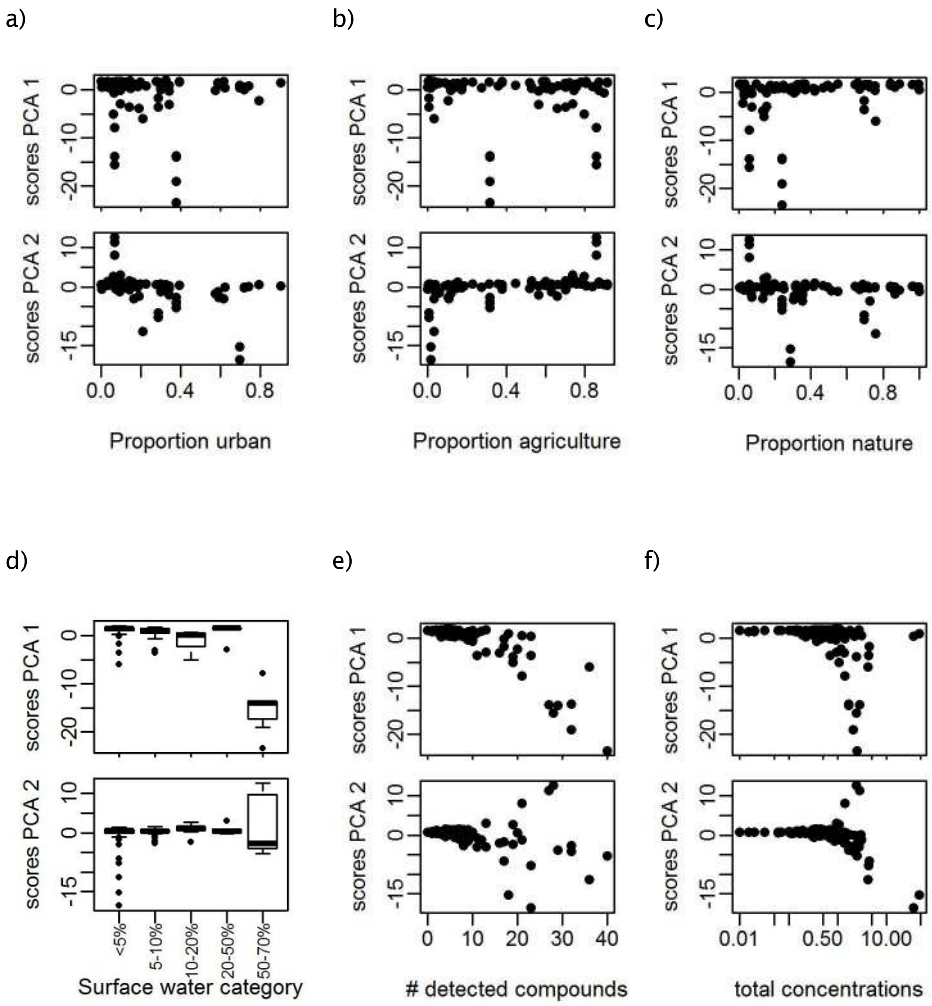


Figure S.1.2.a. Relations between PCA scores of **target** water samples vs a) proportion of urban land-use, b) proportion of agriculture land-use, c) proportion of nature land-use, d) surface water category, e) number of detected chemicals, and f) total concentrations of all chemicals (µg/l).

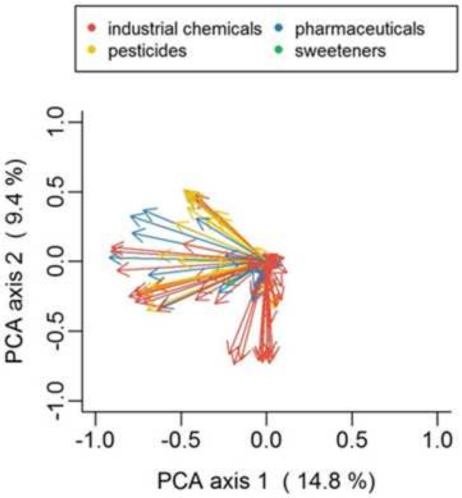


Figure S.1.2.b. Loadings of target chemical in source water on PCA axis 1 and 2. Colors of arrows depict different chemical uses.

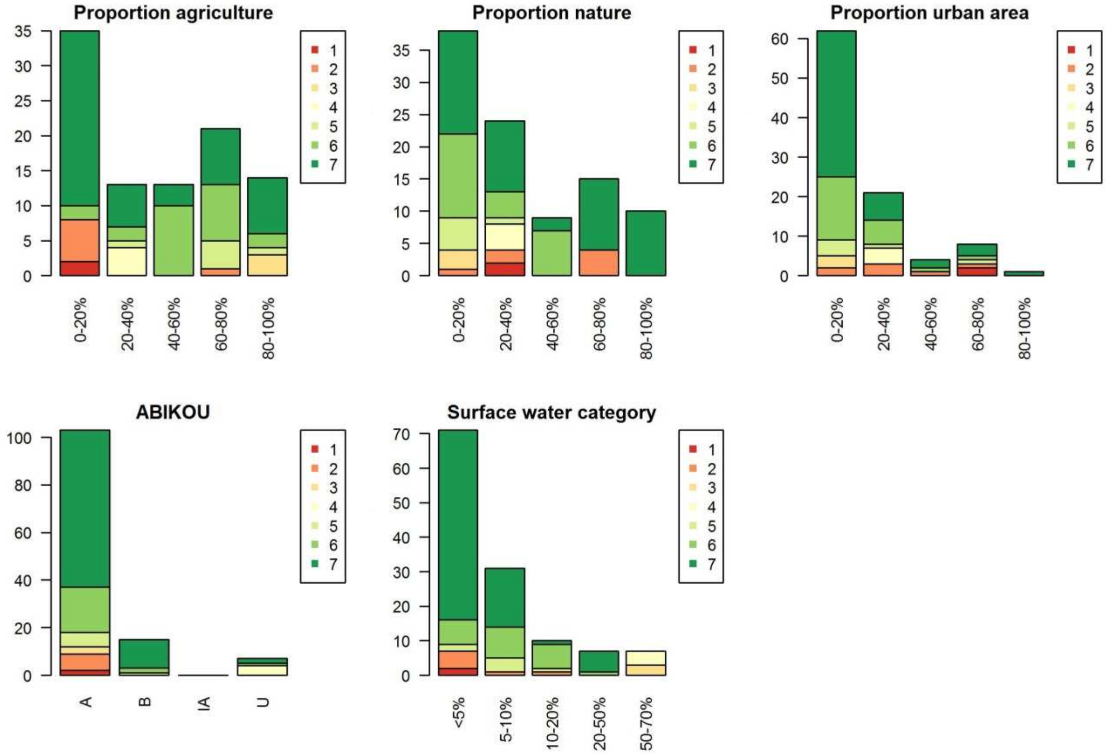


Figure S.I.3.a. Occurrence frequency of 7 clusters of source water samples based on **target** chemical composition, per well typology. The typologies used are: proportion of agricultural land-use, proportion of nature land-use, proportion of urban land-use, ABIKOU class, and influence of surface water infiltration. ABIKOU and categories for surface water infiltration is described in material and method section.

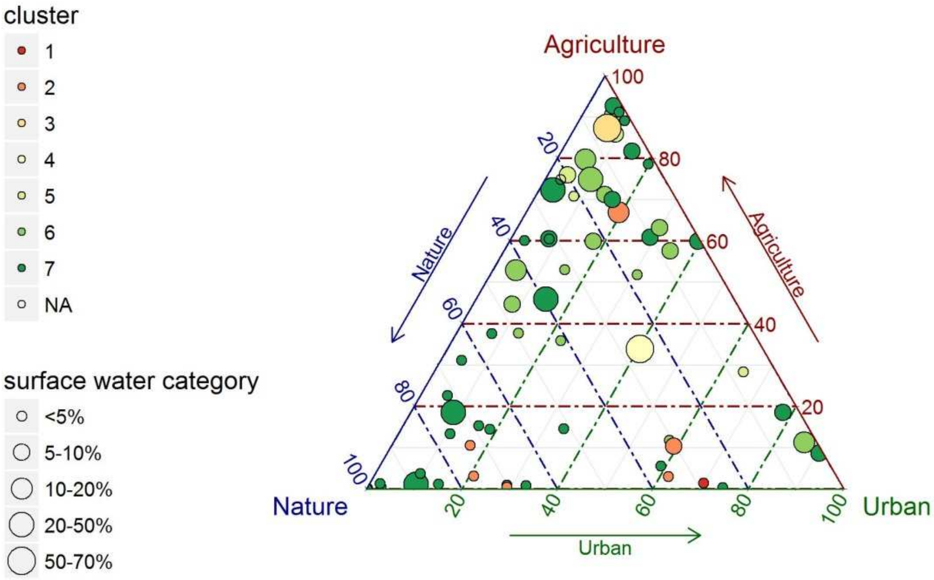


Figure S.I.3.b. Distribution of 7 clusters of source water samples based on **target** chemical composition, projected on a ternary plot of proportion (%) of three different land-uses: agriculture, urban, and nature. Size of the circles depict surface water infiltration category.

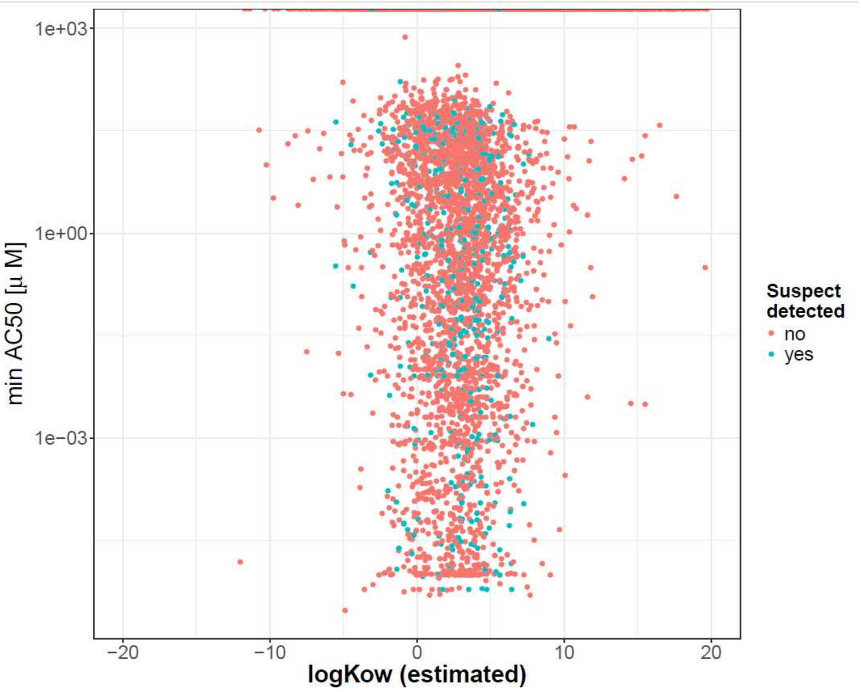


Figure S.I.4. Hydrophobicity and minimum AC50 Toxcast toxicity data compared for detected and non-detected suspects.

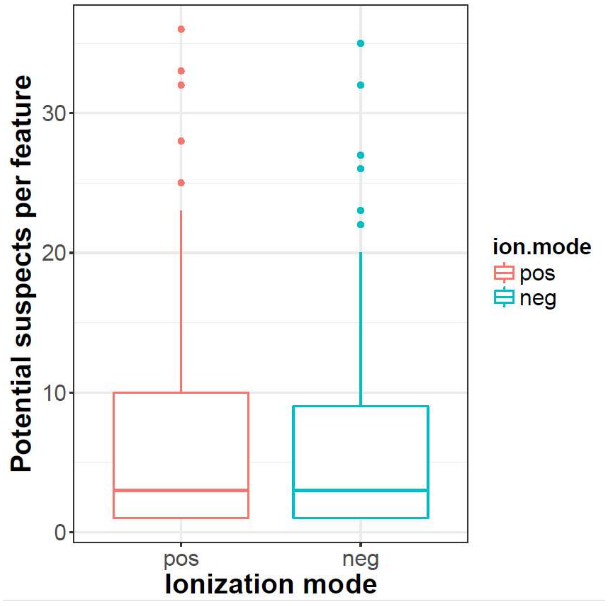
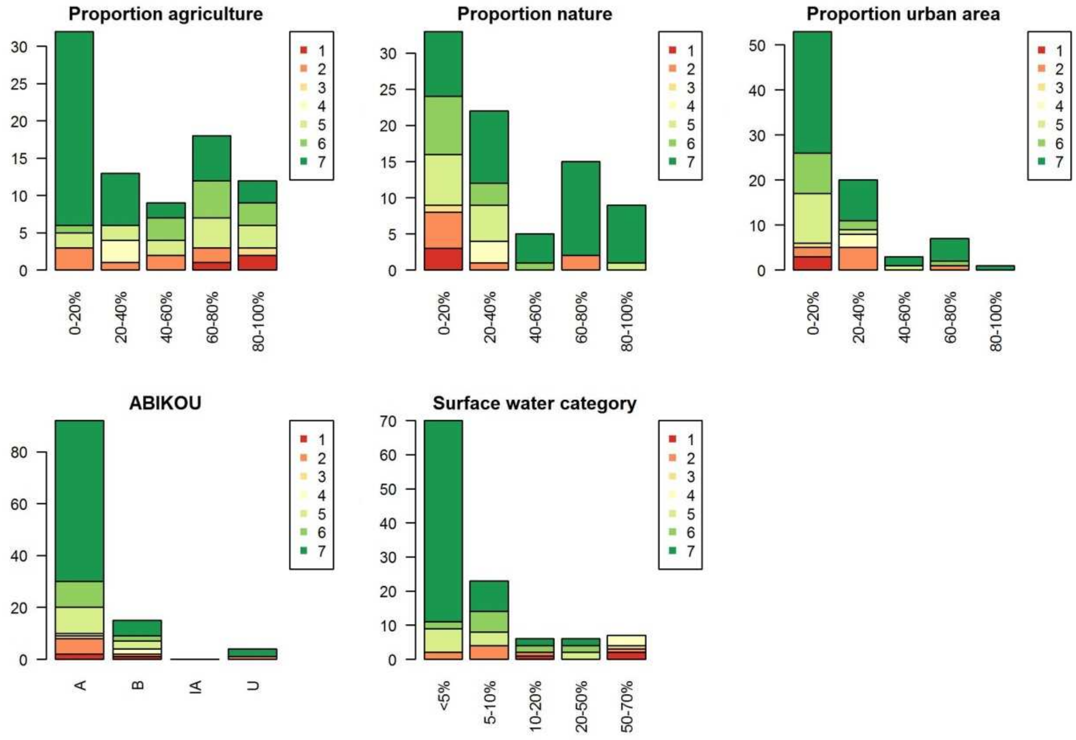
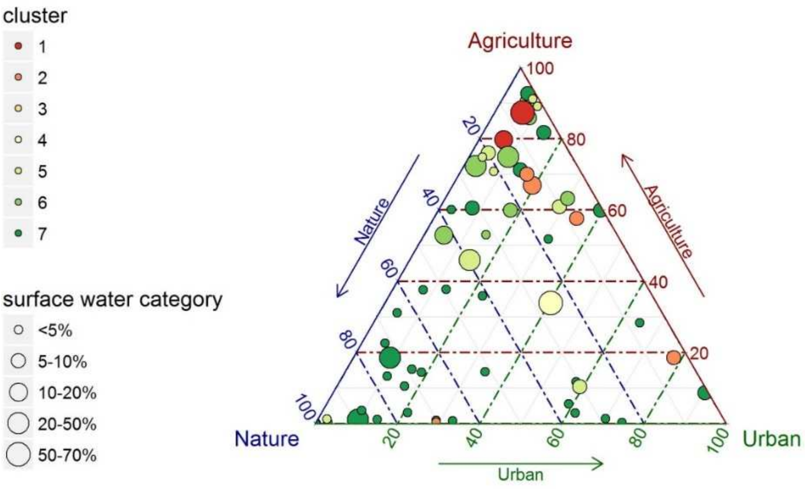


Figure S.I.5. Features match to a maximum of 36 and 35 different **suspects** in positive and negative mode, respectively, median 3 different suspects both in the positive and negative ionisation mode



S.I.6.a. Occurrence frequency of 7 clusters of source water samples based on **suspect** chemical composition, per well typology. The typologies used are: proportion of agricultural land-use, proportion of nature land-use, proportion of urban land-use, ABIKOU class, and influence of surface water infiltration. ABIKOU and categories for surface water infiltration is described in material and method section.



S.I.6.b. Distribution of 7 clusters of source water samples based on **suspect** chemical composition, projected on a ternary plot of proportion (%) of three different land-use: agriculture, urban, and nature. Size of the circles depict surface water category.

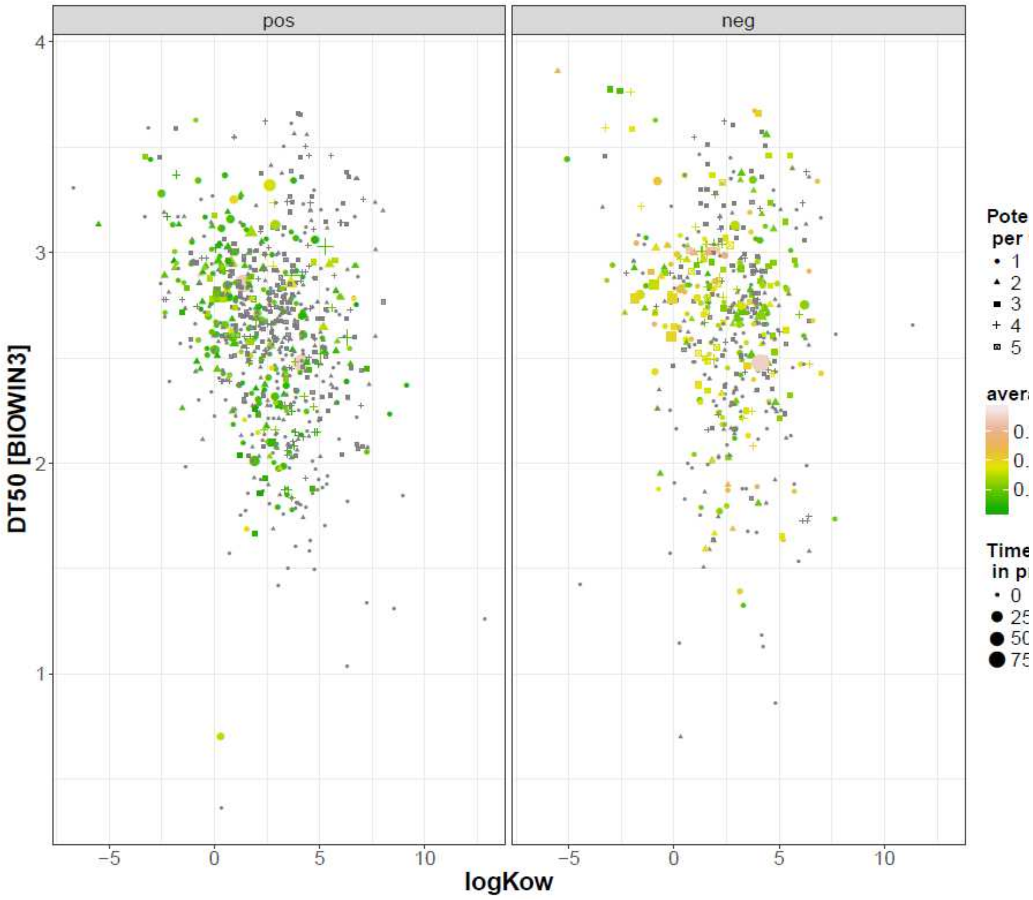


Figure S.I.7. Retrieved suspects in produced drinking water in relation to log Kow and DT50, in grey points suspects that occur in source waters but not in produced waters.

Table S.I.1. Suspect list (details)

ion	ture			Suspect					
+	340.0240 / 25.89	3.390.159	1	1,1,2,2,3,3,4,4-octafluoro-N-(2-hydroxyethyl)-N-methylbutane-1-sulphomide (EINECS)	3.390.175.399	93894-53-2	C7H9F8NO3S	no	Annex III, pre-registration
+	571.0931 / 22.46	5.700.851	4	1-Propaminium, N-(carboxymethyl)-N,N-dimethyl-3-[[(3,3,4,4,5,5,6,6,7,7, 8,8,8-tridecafluorooctyl)sulfonyl]amino]-, inner salt (TSCA)	570.085.795	34455-29-3	C15H19F13N2O4S	yes	no
+	365.0221 / 15.30	3.640.143	2	1-Octanol, 3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluoro- (TSCA, DSL, ENCS, AICS)	3.640.132.816	647-42-7	CF3(CF2)5CH2CH2OH	yes	no
+	511.0994 / 23.83	5.100.914	1	(3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluorooctyl)triethoxysilane	5.100.895.879	51851-37-7	C14H19F13O3Si	yes	Annex III, pre-registration
+	511.0994 / 23.83	5.100.914	1	(3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluorooctyl)diisopropoxyhydroxysilane	5.100.895.879		C14H19F13O3Si	no	no
+	457.9883 / 8.56	4.569.804	9	Glycine, N-ethyl-N-[(undecafluoropentyl)sulfonyl]-, sodium salt (TSCA, NDSL)	4.569.817.834	68555-69-1	C9H7F11NNaO4S	no	Annex III, pre-registration
-	luorobutyric acid (	2.139.865	4	Heptafluorobutyric acid	2.139.864.768	375-22-4	C4HF7O2	yes	Annex III, pre-registration
-	PFOA	4.139.737	4	Perfluorooctanoic acid	4.139.737.026	335-67-1	C8HF15O2	yes	Annex III Inventory
-	485.3014 / 25.75	4.863.076	5	2,2'-ethylidenebis(4,6-di-tert-butylphenyl)fluorophosphonite	4.863.062.955	118337-09-0	C30H44FO2P	yes	Annex III Inventory
-	398.9372 / 41.68	3.999.434	2	Perfluorohexanesulfonic acid	3.999.438.818	355-46-4	C6HF13O3S	no	Annex III Inventory
-	336.8982 / 31.31	3.379.061	1	K-PFBS	3.379.061.505	29420-49-3	K+(C4F9SO3)-	yes	Annex III, pre-registration
-	244.9861 / 20.84	2.459.924	5	2,2,3,3,4,4,5,5-octafluorovaleric acid (EINECS)	2.459.927.051	376-72-7	C5H2F8O2	yes	Annex III, pre-registration
-	346.9758 / 40.73	3.479.825	2	1,1,2,2,3,3,4,4,5,5-decafluoro-1-[(trifluorovinyl)oxy]pentane (English, French) (EINECS)	3.479.819.815	73928-40-2	C7HF13O	no	Annex III, pre-registration
-	346.9758 / 40.73	3.479.825	2	Perfluoroheptanal	3.479.819.815	63967-41-9	C6F13CHO	no	no
-	298.9418 / 23.84	2.999.499	47	Perfluorobutanesulfonic acid	2.999.502.689	375-73-5	C4HF9O3S	no	Annex III Inventory
-	362.9668 / 25.90	3.639.765	20	Hexanoic acid, 2,2,3,3,4,4,5,6,6,6-decafluoro-5-(trifluoromethyl)- (TSCA, NDSL)	3.639.768.961	15899-29-3	C7HF13O2	no	Pre-registration process
-	362.9668 / 25.90	3.639.765	20	Heptanoic acid, tridecafluoro- (TSCA, NDSL, ENCS, AICS)	3.639.768.961	375-85-9	C7HF13O2	yes	Annex III, pre-registration
-	280.9521 / 20.93	2.819.588	4	1,1,2,2,3,3,4,4-octafluorobutane-1-sulphonic acid (EINECS)	2.819.596.907	70259-86-8	C4H2F8O3S	no	Annex III, pre-registration
-	300.9386 / 23.39	3.019.461	2	1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulphonyl fluoride (PBSF)	3.019.459.324	375-72-4	C4F9SO2F	yes	no
-	317.0237 / 18.11	3.180.314	2	2-Propenoic acid, 3,3,4,4,5,5,6,6,6-nofluorohexyl ester (TSCA, DSL, ENCS)	3.180.302.334	52591-27-2	C9H7F9O2	yes	no
-	262.9766 / 19.73	2.639.833	1	Pentanoic acid, nonafluoro- (TSCA, NDSL, ENCS, AICS)	2.639.832.833	2706-90-3	C5HF9O2	yes	Annex III, pre-registration
-	297.9585 / 24.83	2.989.656	1	FBSA	2.989.662.533	30334-69-1	C4F9SO2NH2	no	no
-	641.2101 / 25.76	6.422.149	2	(Perfluorooctyl)hexadecene	6.422.154.307	1244062-16-5	F(CF2)8CHCH(CH2)14H	no	no
-	348.9401 / 25.57	3.499.469	4	Perfluoropentanesulfonic acid	3.499.470.753	2706-91-4	C5HF11O3S	no	Annex III, pre-registration
-	470.0135 / 25.18	4.710.197	1	EtFHxSA	471.017.381	34455-03-3	C6F13SO2N(C2H5)CH2CH2OH	no	Annex III, pre-registration