BTO 2017.011| February 2017

BTO report

Exploring the boundaries of non-target screening with Liquid Chromatography coupled to ESI-MS

KWR Watercycle Research Institute

BTO

Exploring the boundaries of non-target screening with Liquid Chromatography coupled to ESI-MS

BTO 2017.011 | February 2017

Project number
400554-125

Project manager
S.A.E. Kools

Client
BTO

Quality Assurance
W.P de Voogt

Author(s)
J.A. van Leerdam, B.L. Bajema (Vitens), R.M.A. Sjerps, B.A. Wols, P.S. Bäuerlein, T.L. ter Laak, M.M.E. van der Kooi en E. Emke

Sent to
This report is distributed to BTO-participants.
A year after publication it is public.

Year of publishing
2017

More information
Ing. Ton van Leerdam
T 030 60 69 624

ton.van.leerdam@kwrwater.nl

Keywords: mass spectrometry, non-target screening, response factor

PO Box 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

KWR Watercycle Research Institute

BTO | February 2017 © KWR

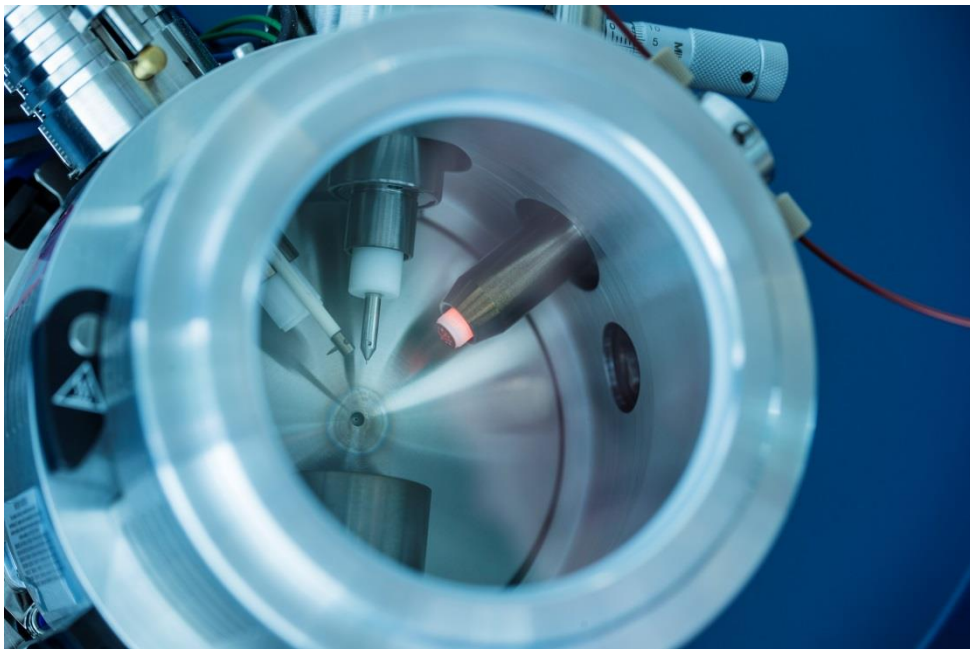
All rights reserved.
No part of this publication may be reproduced, stored in an automatic database, or transmitted, in any form or by any means, be it electronic, mechanical, by photocopying, recording, or in any other manner, without the prior written permission of the publisher.

BTO *Managementsamenvatting*

Hoe breed is brede screening ?

Auteur(s): J.A. van Leerdam, B.L. Bajema (Vitens), R.M.A. Sjerps, B.A. Wols, P.S. Bäuerlein, T.L. ter Laak, M.M.E. van der Kooi en E. Emke

Vitens en KWR hebben op basis van uitgebreide metingen, literatuuronderzoek en ervaring twee modellen ontwikkeld die inzicht bieden in de inzetbaarbeid van non-target screening met vloeistofchromatografie (LC) en hoge-resolutie-massaspectrometrie (MS): LC-HRMS. Uit het onderzoek bleek dat er diverse verbanden bestaan tussen meetgevoeligheid en fysisch-chemische stofeigenschappen zoals elektronegativiteit en het aantal stikstofatomen in het molecuul. Het voorspellen van de meetgevoeligheid is complex omdat daarbij meerdere eigenschappen tegelijkertijd een rol spelen, naast de instrumentele instellingen. Model 1 voorspelt met 80% zekerheid of een stof wél of niet analyseerbaar is met LC-HRMS. Deze betrouwbaarheid wordt voldoende geacht voor een voorspellingsmodel. Model 2 bepaalt in welke van de vier meetgevoeligheidsklassen een stof valt, met 60% tot 76% zekerheid. Dit wordt nog onvoldoende geacht voor een betrouwbare voorspelling. Met model 1 is voor 150 voor drinkwater relevante stoffen vastgesteld in hoeverre ze te analyseren zijn. Van deze stoffen kan volgens het ontwikkelde model 65% worden geanalyseerd met LC-HRMS.



Ionisatiekamer van de Orbitrap massaspectrometer

Belang: zoveel mogelijk chemische stoffen die in het milieu voorkomen detecteren

Wereldwijd worden er steeds meer stoffen geproduceerd. Een deel van deze stoffen komt ook in het milieu terecht en kan daardoor een bedreiging vormen voor de bronnen voor drinkwaterbereiding. Een zo compleet mogelijke chemische screening is noodzakelijk om de verspreiding van stoffen en de eventuele risico's die ze met zich meebrengen goed te kunnen inschatten. Met behulp van een brede screening met de combinatie van vloeistofchromatografie (LC)

en hoge-resolutie-massaspectrometrie (MS) is het mogelijk in één analysegang zowel op bekende als onbekende stoffen te screenen. Deze methode is inmiddels circa tien jaar beschikbaar en wordt wereldwijd steeds meer toegepast. In 2013 hebben KWR en Vitens een begin gemaakt met de verdere harmonisatie van de werkwijze van de methode voor gebruik in de drinkwaterlaboratoria. Over de reikwijdte en de grenzen van deze detectiemethode en de betrouwbaarheid van de kwantificering is echter nog weinig bekend. Kennis hierover is nodig om zo goed mogelijk in beeld te krijgen in hoeverre non-target screening het mogelijk maakt om het steeds groter wordende bereik aan chemische stoffen in beeld te krijgen.

Aanpak: metingen aan een groot aantal doelstoffen combineren met statistische big-data-technieken

In totaal zijn in de laboratoria van Vitens en KWR bijna 500 stoffen geanalyseerd met de brede screeningsmethode, waarbij de meetgevoeligheid is geregistreerd. Op basis van een literatuurstudie en de ruime ervaring met deze techniek zijn chemische eigenschappen geselecteerd die waarschijnlijk van invloed zijn op de ionisatie in de massaspectrometer en daarmee op de meetgevoeligheid van de techniek. Met de nieuwste statistische big-data-technieken zijn relaties opgespoord tussen de gemeten gevoeligheid en in totaal 41 fysisch-chemische eigenschappen uit een set van 1300 vrij beschikbare molecuul descriptoren (PaDEL descriptoren).

Resultaten: maat voor meetgevoeligheid ontwikkeld

Uit de analyse van de ruim 500 stoffen bleek dat de meetgevoeligheid voor het overgrote deel van de stoffen lager is dan de gevoeligheid voor de gebruikelijke interne standaardstoffen. De in het verleden opgegeven concentraties in equivalenten interne standaard (atrazine-d5 en bentazon-d6) zijn daarom te laag, de werkelijke concentraties blijken tot meer dan een factor 10.000 hoger te kunnen liggen. Voor een betere concentratieberekening zijn vier gevoeligheidsklassen gedefinieerd. Op basis van deze klasse-indeling kan een correctie voor de concentratie worden toegepast.

Uit het onderzoek bleek dat er diverse verbanden bestaan tussen meetgevoeligheid en fysisch-chemische stofeigenschappen als electronegativiteit en het aantal stikstofatomen in het molecuul. Het voorspellen van de meetgevoeligheid is complex omdat daarbij meerdere eigenschappen tegelijkertijd een rol spelen, naast de instrumentele instellingen.

Op basis van 32 fysisch-chemische eigenschappen zijn twee modellen ontwikkeld:

- Model 1 voorspelt met 80% zekerheid of een stof wél of niet analyseerbaar is met LC-HRMS. Deze betrouwbaarheid wordt voldoende geacht voor een voorspellingsmodel;
- Model 2 bepaalt in welke van de vier meetgevoeligheidsklassen een stof valt, met 60% tot 76% zekerheid. Dit wordt nog onvoldoende geacht voor een betrouwbare voorspelling.

Met model 1 is voor 150 voor drinkwater relevante stoffen vastgesteld in hoeverre ze te analyseren zijn. Van deze stoffen kan volgens het ontwikkelde model 65% worden geanalyseerd met LC-HRMS.

Het tweede model zal in de toekomst verfijnd moeten worden om de meetgevoeligheid goed te kunnen voorspellen.
De verwachting is dat modellen in de toekomst een steeds grotere rol zullen spelen bij het voorspellen van de analyseerbaarheid én bij het voorspellen van de zuiveringsrendementen en toxiciteit van chemische stoffen.

Implementatie: gegevens gedeeld in Europese databank en workshop brede screening

De massaspectra van een deel van de gemeten stoffen zijn ingevoerd in de Europese databank ‘Massbank’, zodat ook andere instituten van deze gegevens gebruik kunnen maken. In 2017 zullen de resultaten van dit onderzoek samen met resultaten van aanpalende BTO-projecten zoals *Bevestigen suspects* en *Kwaliteitsborging brede screening* in een workshop met de drinkwaterlaboratoria worden gedeeld. In de workshop zal centraal staan in hoeverre de nieuwe inzichten in de praktijk toegepast kunnen worden.

Rapport

Dit onderzoek is beschreven in het rapport *Exploring the boundaries of non-target screening with Liquid Chromatography coupled to ESI-MS* (BTO 2017.011).
Tevens worden de onderzoeksresultaten ingediend bij het wetenschappelijke tijdschrift ‘International Journal of Mass Spectrometry’.

More information
Ton van Leerdam
T 030 60 69 625
E ton.van.leerdam@kwrwater.nl

KWR
PO Box 1072
3430 BB Nieuwegein
The Netherlands

KWR Watercycle Research Institute

Contents

1 Introduction

1.1 Chemicals in the aquatic environment

Worldwide over 100 million organic compounds are registered in the Chemical Abstract Service (CAS) database, approximately 70.000 have been registered for commercial application in Europe (Schwarzenbach, 2006) and 11.000 are produced or used in volumes over 100 tons per year in the European Union (European Chemicals Agency Information on Chemicals).

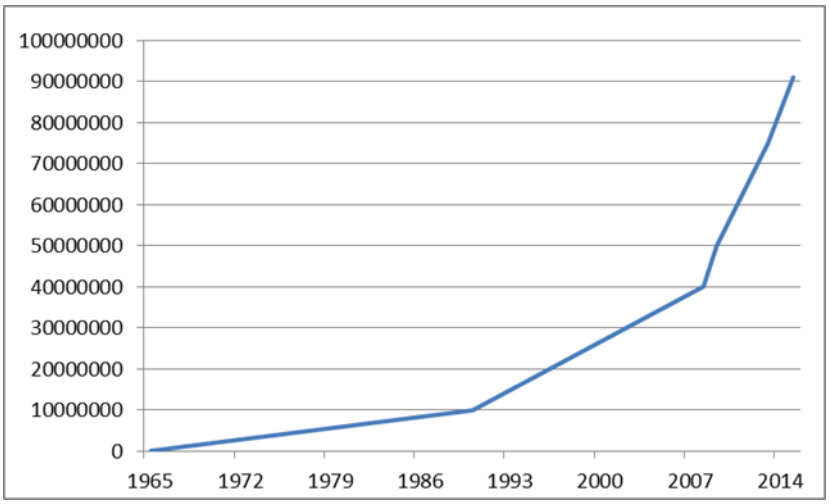


FIGURE 1-1: REGISTERED CHEMICALS IN CAS-DATABASE OVER THE LAST 50 YEARS

The release of compounds and resulting risks for human health and the environment have raised the concern of scientists and policy makers. The number of substances is too large to be monitored or to assess any substance and its risks separately. Therefore, there is a need for an analytical methodology that in a relatively simple way provides reliable information on the presence of anthropogenic substances in water.

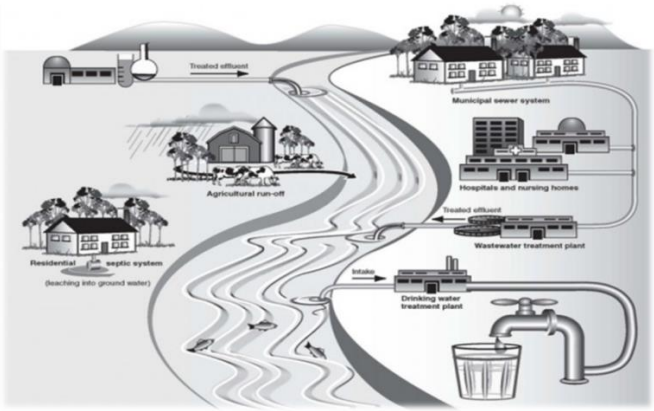


FIGURE 1-2: RELEASE OF CHEMICAL COMPOUNDS TO SURFACE WATER

Drinking water has to meet several water quality standards for inorganic and organic substances. In The Netherlands, a generic target value with an alerting function has been set at 1 µg/L for all organic anthropogenic chemicals in sources of drinking water.

The regulatory framework leads to extensive monitoring programs by the drinking water companies. Different compounds can be detected by different analytical methods, a costly approach. In addition, the targeted approach is narrowed by the pre-selection of target compounds. The liquid chromatography high resolution mass spectrometry (LC-ESI-HR-MS) technique can combine target and non-target analyses, which provides a broader overview of the presence of substances in water (ter Laak et al., 2012).

Despite the detection of a wide range of organic compounds, the analytical conditions are not ideal for all compounds. Reemtsma et al. (2016) highlight the issue of Persistent Mobile Organic Compounds (PMOCs) from an environmental perspective and assess the gaps that appear to exist in terms of analysis, monitoring, water treatment en regulation. In their paper they elaborate on strategies for how to narrow these gaps with the intention to better protect water resources.

The analysis of a chemical compound depends on a lot of variables like isolation recovery, chromatographic behaviour and ionisation efficiency in the mass spectrometer. The goal of the present study is to gain insight in the analytical limitations of the LC-HRMS method for the detection of compounds relevant for drinking water when used as a tool for determining the water quality.

An LC-HRMS screening was performed for a reference set of 223 compounds by KWR and 294 compounds by Vitens. The dataset is used to study the sensitivity of the LC-HRMS method in more detail and to predict the response factor (relative to the internal standard) using physicochemical compound properties.

1.2 Goal of the study

Using experimental data from LC-HRMS screening coupled to ESI, we aim to get a better understanding of the limitations and the sensitivity of the LC-HRMS method. The goal of this study is threefold:

- What are the limits of the LC-HRMS screening method? Which chemicals or chemical groups can be detected with the method and which ones cannot?
- Which physicochemical properties can be identified that affect the detectability of the chemicals with LC-HRMS? For example polarity, volatility and molecular size.
- Can the sensitivity of the detection be predicted using one or a combination of multiple physicochemical properties?

2 Non-target screening with LC-HRMS

This chapter provides some background information on non-target LC-HRMS screening and does not purport to be complete. After sample collection (paragraph 2.2), sample preparation (paragraph 2.3), separation process during chromatography (paragraph 2.4) and the detection of compounds (paragraph 2.5), results are presented of a literature study on physicochemical properties that mainly influence ionisation efficiency (paragraph 2.6).

2.1 Introduction

Gas chromatography (GC) and Liquid chromatography (LC) coupled to a Mass Spectrometry (MS) are the most frequently applied techniques to detect and identify organic substances in environmental chemistry (Schwarzenbach et al., 2006). Often MS is complemented with UV/Vis spectroscopy to get additional information on the compounds. The reason why LC-HRMS is the most common technique is its versatility. This technique covers the largest range of organic compounds, whereas GC-MS is most commonly used for uncharged, small and reasonably volatile compounds (Figure 2-1). To impart volatility to otherwise non-volatile compounds, derivatisation in GC-MS is frequently used.

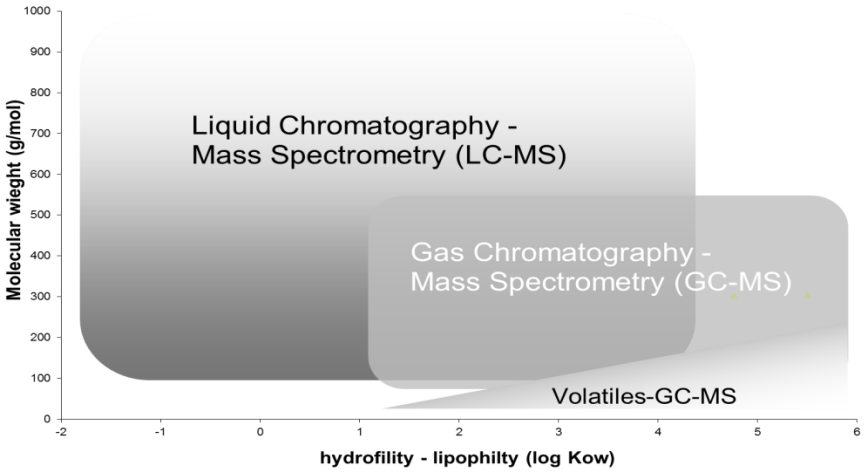


FIGURE 2-1: DOMAINS OF GC-MS AND LC-HRMS TECHNIQUES IN THE TWO DIMENSIONS OF MOLECULAR WEIGHT AND HYDROPHILITY (SOURCE: KWR REPORT 08.013)

In this chapter the most common LC and MS techniques for non-target screening will be explained. Furthermore the whole procedure from sample collection to data analysis will be explained as the choice of technique in each step has a significant impact on the analysis. Figure 2-2 shows the pathway from sampling to data analysis and in Table 2-1 selected techniques for each step are presented.

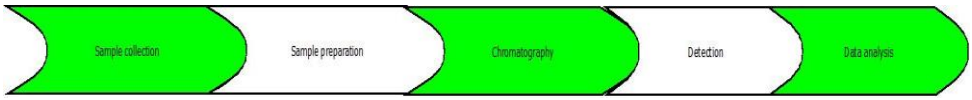


FIGURE 2-2: FLOW CHART OF AN ANALYTICAL PROCEDURE .

TABLE 2-1: DIFFERENT ELEMENTS IN THE NON-TARGET LC-HRMS SCREENING METHOD.

Sample Collection*	Sample Preparation*	Chromatography	Detection
Grab sampling	Liquid-Liquid Extraction (LLE)	Reversed Phase	Electrospray Ionisation (ESI)
Passive sampling	Solid Phase Extraction (SPE)	Normal Phase	Atmospheric Pressure Chemical Ionisation (APCI)
Time integrated sampling	Direct water injection	HILIC	Atmospheric Pressure Photo Ionisation (APPI)
	Evaporation		

* these parts of the LC-HRMS screening are not involved in this study due to the use of pure reference standards without sample collection and sample preparation.

2.2 Sample collection and conservation

The sample is chosen in such a way that it represents the presence of one or multiple analytes in the water. The water sample encompasses the spatial dimensions as well as a temporal component. In both cases the quality of the sample depends upon how accurately the body of water is represented in the laboratory sample.

Sampling is the first step in a multistep process (the analytical procedure) toward receiving meaningful results (see figure). Sampling is also the most critical step in the whole analytical procedure. If the sampling is not carried out cautiously and precisely, the following steps will turn out to be a meaningless exercise. Several things have to be kept in mind. Liquid samples often require the immediate addition of analyte-specific preservatives (e.g. NaN_3 to prevent microbial degradation). For trace-level analyses the sample collection vessel must be composed of a material that will not interfere with the sample (e.g. adsorption of compounds of interest to the bottle wall) and it must be rigorously cleaned before use. The storage temperature is crucial. May the sample be frozen? Will compounds degrade above certain temperatures? To be able to answer all these questions, the sampling steps have to be evaluated and validated. Apart from validating the sampling steps, also recording information of the sample and sampling location is essential (e.g. weather conditions, temperature etc.).

2.3 Sample preparation

After the sample has been taken, the next step in the laboratory is to decide whether the sample needs further treatment. Sometimes it is necessary to remove interferences, perhaps even to completely isolate the intended analyte from its sample matrix. Also further concentration of the analyte can be necessary. Several techniques are available to reach

these goals. To concentrate the analyte liquid/liquid extraction, solid phase extraction (SPE) and evaporation are common techniques. Filtration, liquid/liquid extraction and SPE can be used to separate the analyte from the sample matrix. Also, all these steps must be validated beforehand to make sure they neither result in losses of the analyte, nor contaminate the sample.

2.4 Chromatography

After the sample has been prepared for analysis, it will be subjected to liquid chromatography. Chromatography is a technique to separate and makes use of the distribution of a compound between two different phases, a mobile phase and a stationary phase. In LC the mobile phase is a liquid and the stationary phase can be either liquid or solid. The most common approach to realise such a system is a column (a tube) in which the stationary phase is contained and the mobile phase flows through the interstitial channels. For the separation mechanism to work the two phases must have different properties. The most frequently applied technique for analysis of organic compounds is reverse phase high performance liquid chromatography (RP-HPLC). In this technique the mobile phase is a polar solvent or mixture, whereas the stationary phase is predominantly apolar. In that case the more polar compounds favour the liquid phase (mobile phase) and will move faster through the column. As a consequence they get separated from the more apolar compounds and elute first.

Another chromatographic technique is called HILIC (Hydrophilic Interaction Liquid Chromatography). Also in this case a column is filled with a solid carrier material. However, water molecules on the surface of this solid form a stationary layer which is in contact with the moving liquid layer. This technique is used for the separation of extremely polar compounds (Speksnijder et. al (2012), Vughs et al. (2015)).

Based on the polarity of a chemical one has to decide which of the two techniques should be applied. However, it needs to be mentioned that HILIC is still a technique under development and is at large more difficult to apply than RP-HPLC. In both cases, when the compounds emerge from the column, they still need to be detected and preferably unequivocally identified. Several methods of detection are possible. The most common ones are UV-Vis spectroscopy and mass spectrometry.

2.5 Detection

2.5.1 UV-Vis spectroscopy

UV-Vis spectroscopy or just UV-Vis capitalises on the fact the molecules with π-bonds adsorb light between 200 and 700 nm. The amount and wavelength of light absorbed by a molecule provides useful information on the molecular structure and the quantity. As most chemicals that are interesting for environmental chemists contain double or triple bonds, this technique is a powerful tool for screening environmental samples. HPLC-UV screening is used for more than 10 years by KWR, WML, Evides, Aqualab Zuid and Rijkswaterstaat for the quick and relatively inexpensive monitoring of organic micropollutants in the river Meuse.

Due to the limited identification power of UV, additional identification techniques such as mass spectrometry or Nuclear Magnetic Resonance are commonly used for the identification of unknowns detected by HPLC-UV.

2.5.2 Mass spectrometry

Mass spectrometry is an analytical technique by which chemical substances are identified by the sorting of gaseous ions in electric and magnetic fields according to their mass-to-charge ratios. This means in this case one gets information on the molecular mass of the analyte.

In a mass spectrometer molecules or fragments of molecules are separated according to their mass. Four steps in this process can be distinguished: (i) transfer of the molecules in the sample to the gaseous form, (iI) ionisation in the ionisation chamber, (iii) separation in the mass analyser and (iv) detection. This applies for all types of mass spectrometers. The second step, the ionisation, is the crucial step in mass spectrometry. If a molecule is not or not sufficiently ionised, it is either not detected at all or the signal is weak. The latter will result in a low sensitivity and can result in an underestimation of the actual concentration of a compound in a sample.

There are numerous ways to ionise molecules. The most common ones for environmental analysis are: ESI (electrospray ionisation), APCI (atmospheric pressure chemical ionisation) and APPI (atmospheric pressure photo ionisation), they all differ in the way the molecules are converted into ions and also affect significantly the grade of fragmentation of the molecules. In Figure 2 3 the increasing interest for these 3 ionisation techniques are illustrated.

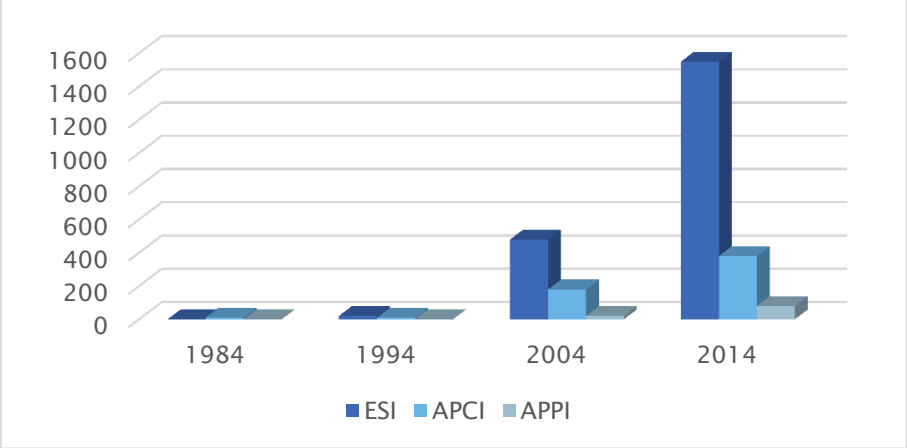


FIGURE 2-3: THE AMOUNT OF SCIENTIFIC PAPERS REGARDING THE USE OF ESI, APCI AND APPI IN COMBINATION WITH LC-HRMS FOR ENVIRONMENTAL ANALYSIS IN THE LAST 40 YEARS (SOURCE: SCOPUS)

Electrospray ionisation (ESI) is a technique used in mass spectrometry to produce ions using an electrospray in which a high voltage is applied to a liquid to create an aerosol. The first use of ESI in combination with mass spectrometry was reported by Malcolm Dole in 1968. John Bennett Fenn was awarded the 2002 Nobel Prize in Chemistry for the development of electrospray ionisation mass spectrometry in the late 1980s.

ESI is a so-called 'soft ionisation' technique, since there is very little fragmentation. In this study, ESI is used for all experiments. In combination with LC, ESI is the most frequently applied ionisation technique due to its versatility. A wide range of compounds from very hydrophilic (log Kow = -2) to weakly hydrophilic (log Kow = + 3.5) can be analysed. Also the molecular mass range is extensive (50 to 2000 Da). This is illustrated in Figure 2.4 where the global domains of the three ionisation techniques are shown in the two dimensions of molecular weight and hydrophobicity.

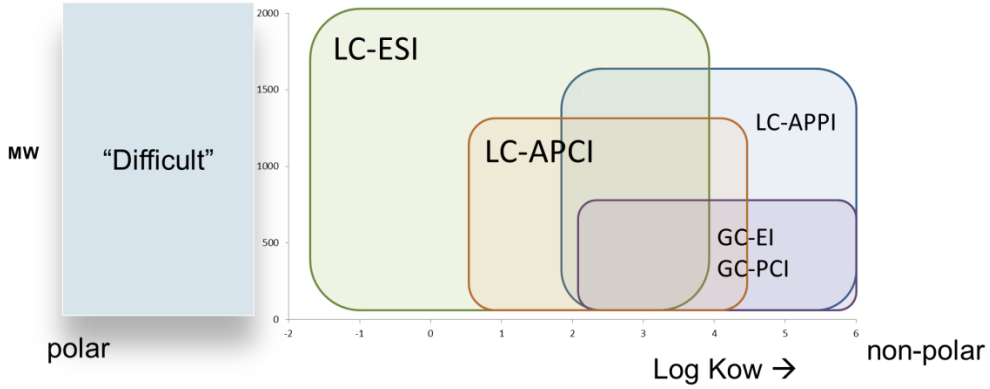


FIGURE 2-4: GLOBAL DOMAINS OF FOUR DIFFERENT SIONISATION TECHNIQUES IN THE TWO DIMENSIONS OF MOLECULAR IWEIGHT AND HYDROPHOBICITY.

The liquid containing the analyte(s) of interest is introduced into the ionisation chamber through a capillary with an electric potential difference (Figure 2-5) and dispersed by electrospray into a fine aerosol. Because the ion formation involves extensive solvent evaporation (also termed desolvation), the typical solvents for electrospray ionisation are prepared by mixing water with volatile organic compounds (e.g. methanol or acetonitrile). To decrease the initial droplet size, compounds that increase the conductivity (e.g. acetic acid) are customarily added to the solution. These species also act to provide a source of protons to facilitate the ionisation process.

The ions observed by mass spectrometry may be quasimolecular ions created by the addition of a hydrogen cation and denoted $[M + H]^+$, or of another cation such as sodium ion, $[M + Na]^+$, or the removal of a hydrogen nucleus, $[M - H]^-$. Multiply charged ions such as $[M + nH]^{n+}$ are often observed. For LC-HRMS screening, the positive as well as the negative ionisation mode must be applied to be able to detect all compounds.

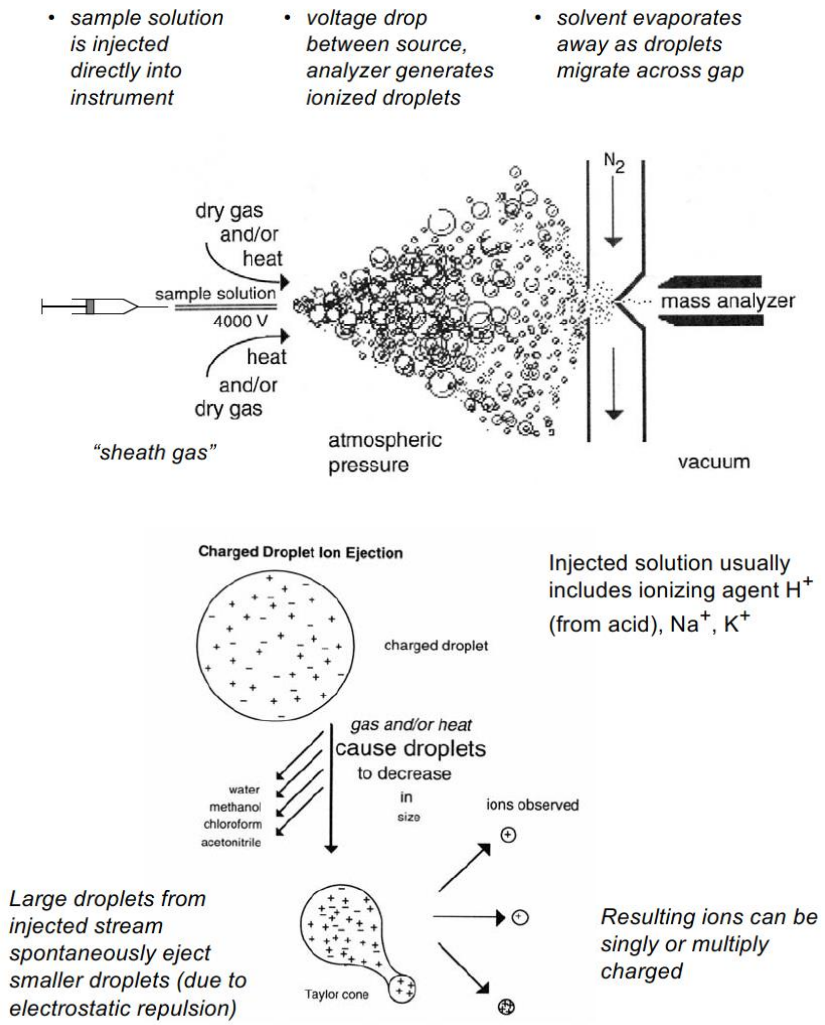


FIGURE 2-5: PRINCIPLE OF ELECTROSPRAY IONISATION (ESI).

There are many types of mass spectrometers using magnetic or electric fields, each type which its own strengths and weaknesses.

- A sector field mass analyser uses an electric and/or magnetic field to affect the path and/or velocity of the charged particles in some way.
- The time-of-flight (TOF) analyser uses an electric field to structural accelerate the ions through the same potential, and then measures the time they take to reach the detector.
- In an ion trap, ions are trapped and sequentially ejected

Beside this, there are several important analyser characteristics. The mass resolving power is the measure of the ability to distinguish two peaks of slightly different m/z. Table 2-2 shows the importance of this feature. Without the accurate mass of the molecules, three completely different compounds with a molecular weight of 84 would be indistinguishable by MS with a low resolving power.

TABLE 2-2 THREE DIFFERENT COMPOUNDS WITH THE SAME NOMINAL MASS BUT DIFFERENT ACCURATE MASSES.

Elemental composition	C_6H_{12}	C_5H_8O	$C_4H_8N_2$
Nominal mass	84.1	84.1	84.1
Accurate mass	84.0939	84.0575	84.0688

There are several ways to achieve accurate masses, the two most common techniques are ion traps (e.g. Orbitrap) and time of flight (e.g. QTOF) in combination with high resolution (HR) MS. This means the determination of mass values accurately up to more than one decimal places, thereby enabling distinguishing different structural formulas having the same nominal mass.

Orbitrap

In an Orbitrap HRMS the ions are trapped in chamber between an inner and an outer electrode. There they oscillate along the inner electrode (Figure 2-6).

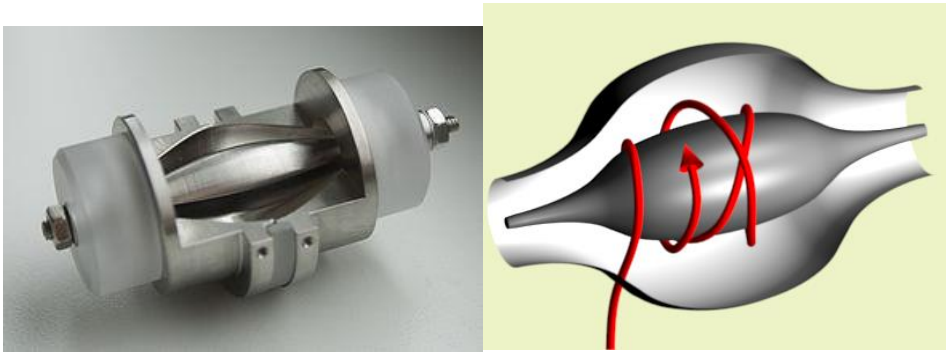


FIGURE 2-6: PICTURE OF THE ORBITRAP CHAMBER (LEFT) AND THE TRAJECTORY OF AN ION IN THE ORBITRAP CHAMBER (RIGHT).

The frequency of these harmonic oscillations is independent of the ion velocity but exclusively inversely proportional to the square root of the mass-to-charge ratio (m/z). Measuring the oscillation and analysing it using Fourier transformation hence gives the accurate mass of the ions. Orbitraps have a high mass accuracy (< 5 ppm), a high resolving power and a high dynamic range.

Time-of-Flight

Time-of-flight mass spectrometry (TOFMS) is a method of mass spectrometry in which an ion's mass-to-charge ratio is determined via a time measurement. Ions are accelerated by an electric field of known strength. This acceleration results in an ion having the same kinetic energy as any other ion that has the same charge. The velocity of the ion depends on the mass-to-charge ratio (heavier ions of the same charge reach lower speeds, although ions with higher charge will also increase in velocity). The time that it subsequently takes for the ion to reach a detector at a known distance is measured. This time will depend on the velocity of the ion, and therefore is a measure of its mass-to-charge ratio. From this ratio and known experimental parameters, one can identify the ion. As with an Orbitrap, a Time-of-

flight instrument have a high mass accuracy (< 5 ppm), a high resolving power and a high dynamic range and therefore used for accurate mass measurements and useful for the identification of unknown compounds.

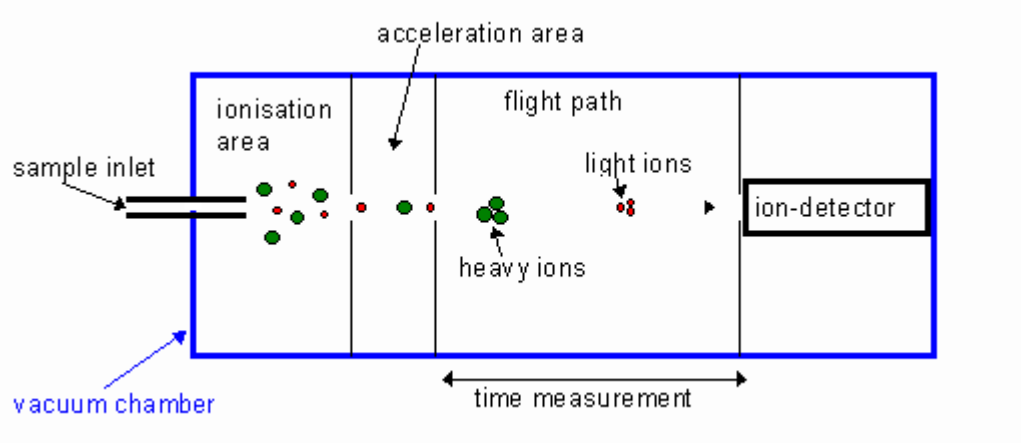


FIGURE 2-7: OVERWIEW OF A TIME-OF-FLIGHT INSTRUMENT.

2.6 Compound variability of the different LC-HRMS stages

A lot of factors affect the degree of how sensitively individual compounds can be detected. These factors can be divided into three groups, related to stage of the LC-HRMS analysis: (i) sample preparation, (ii) chromatography and (iii) ionisation (detection). See also Figure 2.3 for the Flow Chart of an analytical procedure.

While the LC-HRMS non-target screening is used for the detection of a broad range of organic compounds, the analytical conditions are not optimal for all compounds. Sjerps et al. (2016) showed that the response factors of 53 reference compounds detected in the positive ionisation mode, varied within 4 orders of magnitude and for 80% of these compounds the variation remained within 2 orders of magnitude.

2.6.1 Sample preparation

Besides the concentration of the individual compounds in the water sample, sample preparation steps such as extraction (e.g. Solid Phase Extraction and Liquid-Liquid extraction) influence the recovery of compounds.

Experiments at the KWR laboratory with the LC-HRMS screening method (Hogenboom et al. 2009) in drinking water spiked with 100 reference compounds selected for their relevance for surface water quality (Huijzer 2006, Jansen 2009) showed variable recoveries. For the complete list of compounds see Attachment I. A selection of these compounds is also used in the study for the correlation between physicochemical properties and the response factor.

32 of the 100 compounds studied by Jansen are not detectable by LC-HRMS due to a low recovery and/or a low sensitivity. The recoveries of the other 68 compounds were classified into four categories. This classification is presented in table 2-2. Almost half of the compounds (43%) have an acceptable recovery (75-125%). A little less compounds (40%) have a mediate recovery (10-74%). Only a small number of the compounds (15%) have a very low recovery (<10%) or a very high recovery (>125%).

TABLE 2-2 NUMBER OF COMPOUNDS CLASSIFIED IN RECOVERY CLASSES (JANSEN 2009).

Recovery (%)	Nr of compounds	Percentage (%)
<10	8 #	11
10-74	28	40
75-125	30	43
>125	2	3
Sum	68	100

Recovery is 0 %

2.6.2 Matrix effects

To detect chemical compounds by LC-HRMS, the compounds must be separated by LC. In the case that compounds are too polar for separation (e.g. in the case that there is no physical interaction between the chemical compound and the stationary phase) the compound will elute unretained from the LC-column together with a lot of other polar chemical compounds such as humic acids naturally present in water. This mixture of compounds usually causes signal suppression. It is a common LC-HRMS problem and, therefore, should be evaluated with each LC-HRMS method. Evaluation of these effects provides valuable information about the quality of the LC-HRMS method.

2.6.3 Ionisation efficiency

Ionisation efficiency is defined as the ratio of the number of ions generated to the number of molecules consumed in the ion source of a mass spectrometer. Ionisation efficiency is influenced by a multitude of complex processes taking place during electrospray to form a gaseous ion (see paragraph 2.5). The physicochemical properties that affect the ionisation efficiency have been sparsely studied. An overview from the literature is presented below.

The relationships of pK_a and pK_b, the logarithmic acid and base dissociation constants, with ionisation efficiencies were described by four studies (Kruve et al. 2014, Ehrmann et al. 2008, Oss et al. 2010, Leito et al. 2008). The basicity is the ability to become protonated and become a cation. Kruve et al. (2014) observed a correlation between log IE (ionisation efficiency) and pK_a, which indicates that stronger acids (lower pK_a values) tend to have higher ionisation efficiencies.

The extent of charge delocalization in an anion (negative ionisation) can be quantitatively expressed via the WAPS (weighted average positive sigma) parameter (Kruve et al. 2014). The WAPS is calculated with the software COSMO-RS. The smaller the WAPS value, the more delocalized the charge in the anion. Kruve et al. observed higher ionisation efficiency for ions with more efficient charge delocalisation (lower WAPS).

COSMO-RS

A database of 1892 compounds (solvents, small molecules) facilitates instantaneous predictions of log P, solubilities, and other properties. It is easy to add other molecules to the database with a prescribed ADF calculation. Tutorials show step-by-step how to set up COSMO-RS property calculations with the GUI. Scripting tools enable rapid solvent screening, e.g. to find the solvent combination which best partitions a drug and its main contaminant.

The importance of the charge density of the anion for ionisation is explained by two factors: the charge-to-charge repulsion and the solvation energy. The charge-to-charge repulsion is occurring on the surface of the ESI droplet and is responsible for "ion evaporation" from the droplets. On the other hand, lower WAPS also indicate lower solvation energy and lower tendency for ion pairing. As an ion needs to "escape" from the droplet to be detected in MS, the solvation energy describes how much energy is needed to overcome the attractive forces between ion and the solvent molecules.

In addition, Kruve et al. (2014) observed a correlation (R^2 = 0.59) between ionisation efficiency and the number of halogen atoms in the molecule n(Hal). However, also other parameters correlate with the number of halogen atoms (molecular volume was found to be correlated in a positive and WAPS in negative way).

Several studies related hydrophobicity to ionisation efficiency (Cech and Enke 2000, Espinosa et al. 2015, Chalcraft et al. 2009, Henriksen et al. 2005). This parameter is characteristic for the affinity of the protonated forms toward the drops surface. A more hydrophobic compound will have an enhanced affinity for the surface of the droplets and consequently higher ionisation efficiency. However, the hydrophobicity was found not to be statistically significantly related to ionisation efficiency (Oss et al. 2010).

Eight studies observed a relation between the molecular size and the ionisation efficiency. Molecular size can be characterized by molecular weight, molecular volume and molecular surface area. Generally, the larger the molecule the better stabilised its protonated form in the gaseous phase. Bigger molecules tend to have a higher ionisation efficiency (Kruve et al. 2014). Probably, ions formed from larger molecules are more favoured on the surface of the droplet, thus favouring their transfer to the gaseous phase (Kruve et al. 2014, Espinosa et al. 2015). Espinosa et al. (2015) observed an increased response factor in relation to an increased molecular weight inside a family of compounds. Hogan Jr and Fernandez de la Mora (2009) found that for more compact ions, the solvation energy (ΔG) of evaporating ions was found to be higher, therefore delaying the ion evaporation. In addition, Nguyen et al. (2013) observed a positive correlation between ESI response and adjusted mass of the ion (expressed as n(H)/n(C) x molecular mass). Chalcraft et al. (2009) and Oss et al. (2010) both observed that the ESI signal can be predicted using molecular volume.

A total of 14 physicochemical properties that possibly affect the ionisation efficiency were found in literature. These physicochemical properties can be clustered into properties indicating basicity, hydrophobicity and molecular size. Logarithmic transformed values are often used to handle the variation in the values. Based on our experience with LC-HRMS, five additional physicochemical properties: (i) number of N, O, S, P and Cl, (ii) ionisation potential, (iii) electronegativity, (iv) dipole moment and (v) proton affinity, are selected to have a possibly relation with the ionisation efficiency. These parameters are labelled as 'expert judgement' In Table 2.3.

TABLE 2-3 PHYSICOCHEMICAL PROPERTIES IDENTIFIED RELATED TO IONISATION EFFICIENCY.

Parameter	Description	Ionisation mode	Reference
pKa	Measure for acidity in solution	Pos	Oss et al., 2010
		Neg	Kruve et al., 2014
		Pos	Leito et al., 2008

pKb	Measure for basicity in solution	Pos	Ehramm et al., 2008
WAPS	Delocalized charge	Neg	Kruve et al., 2014
log P	Distribution coefficient between polar and non-polar medium. Measure for hydrophobicity.	Neg	Henriksen et al., 2005
log D/logKow	Octanol-water distribution coefficient	Pos	Chalcraft et al., 2009
Hydrofobicity	Physical property repelled from a mass of water	Pos Pos + Neg	Cech and Enke, 2000 Espinosa et al., 2015
Absolute mobility (u0)	-	Pos	Chalcraft et al., 2009
Molecular weight (mw)	Weight of the molecule	Pos + Neg	Espinosa et al., 2015
Adjusted mass (M)	n(H)/n© x molecular mass	Pos	Nguyen et al., 2013
Polar surface area (PSA)	The area of the molecule where hydrogen atoms can attach to the surface	Pos + Neg	Hogan et al., 2009
Length of the alkyl chain (nC)	Molecular structure	Neg	Huffman et al., 2012
Molecular size	Molecular volume of the molecule	Pos Pos	Oss et al., 2010 Chalcraft et al., 2009
Number of acid groups (nCOOH)	Represent the number of acid (COOH) groups in the molecule	Neg	Hellmuth C. et al. 2012
Number of Double Bond Equivalents (DBE)	Measure for aromaticity of organic compounds. Can be calculated from the elemental composition.	Pos + neg	Ghosh B. et al (2015)
Number of N, O, S, P and Cl (nN, nO, nS, nP and nCl)	Represent the number of nitrogen, oxygen, sulphur, phosphor and chlorine atoms	Pos + neg	Expert judgement

	present in the molecule		
Ionisation Potential (IP)	The calculated ionisation energy of the molecule to be ionised. Different calculation methods are available.	Pos + neg	Expert judgement
Electronegativity (e.g. Mpe)	The tendency to capture an electron in a gaseous state	Pos + neg	Expert judgement
Dipole moment (DIPOLE)	Measure for distribution of the positive and negative charges on the molecule	Pos + neg	Expert judgement
Proton affinity (PA)	Measure of gas-phase basicity.	Pos + neg	Expert judgement

In addition to psysicochemical properties, other variables during the ionisation process affect the instrumental sensitivity of the individual compounds. Examples are pH, the percentage organic modifier (e.g. acetonitrile and methanol) during the LC-analysis and experimental settings like cone voltage and capillary temperature of the electrospray ionisation.

The presence of specific chemical groups (e.g. nitrogen, hydroxyl) can also affect the sensitivity of the compounds. This factor is partly incorporated in the identified physicochemical properties by properties such as the length of the alkyl chain and the number of acid groups. In the SMILES notation for chemical compounds, the specific chemical groups are stored in an encrypted form. It is recommended to include the information from the SMILES notation in a possible follow-up research.

SMILES notation

The simplified molecular-input line-entry system (SMILES) is a specification in form of a line notation for describing the structure of chemical species using short ASCII strings. SMILES strings can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules.

The original SMILES specification was initiated in the 1980s. It has since been modified and extended. In 2007, an open standard called "OpenSMILES" was developed in the open-source chemistry community. Other 'linear' notations include the Wiswesser Line Notation (WLN), ROSDAL and SLN.

2.7 Existing QSAR models

A QSAR model is a regression or classification model used in the chemical and biological sciences and in engineering. Like other regression models, QSAR regression models relate a set of "predictor" variables (X) to the potency of the response variable (Y), while classification

QSAR models relate the predictor variables to a categorical value of the response variable. A literature research for models to predict the sensitivity of compounds analysed by LC-HR-MS results in some scarce information. Four models/databases were studied in more detail.

1. https://ochem.eu
   This is an on-line chemical database with properties of chemical compounds such as melting point, water solubility and logD. OCHEM contains 1768810 experimental records for about 516 properties collected from 12438 sources. Based on the experimental data published in the OCHEM database, QSAR models for predictions of chemical properties for a set of compounds can be applied. Beside the chemical information, this database contains information and libraries about structural alerts for endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.
2. https://eurl-ecvam.jrc.ec.europa.eu/databases/jrc-qsar-model-database
   The Joint Research Centre (JRC) QSAR Model Inventory is an inventory of information on the validity of (Q)SAR models that have been submitted to the JRC. The database is intended to help to identify valid (Q)SARs, e.g. for the purposes of REACH. The QSAR Model Reporting Format (QMRF) is a harmonised template for summarising and reporting key information on (Q)SAR models, including the results of any validation studies. The information is structured according to the OECD principles for the validation of (Q)SAR models.
3. http://qsardb.org/repository
   QsarDB hosted by the Molecular Technology Group of the University of Tartu in Estonia is developing and operates domain-specific digital data exchange standards and tools that enables research groups, project teams and institutions to share and present Quantitative Structure-Activity Relationships (data and models). The QsarDB repository is designed for models produced with all statistical and mathematical algorithms that qualitatively or quantitatively express the relationship between the chemical structure and the responses of a compound. This information includes chemico-biological activity (QSAR), physicochemical properties (QSPR), toxicity (QSTR), metabolism (QSMR), reactivity (QSRR), retention (QSRR), permeability (QSPR), pharmacokinetics (QSPR), bioavailability (QSBR), binding (QSBR), etc.
4. http://qsar.food.dtu.dk
   This Danish (Q)SAR Database includes estimates from more than 200 (Q)SARs from free and commercial platforms and related to physicochemical properties, ecotoxicity, environmental fate, ADME and toxicity. (Q)SAR predictions for more than 600,000 chemical substances can be searched, sorting can be made on chemical similarity, and profiles for individual substances can be downloaded. The database is developed by the National Food Institute, Technical University of Denmark, with support from the Danish Environmental Protection Agency, the Nordic Council of Ministers and the European Chemicals Agency.

These four databases particularly contain information about chemical properties and QSAR models to predict toxicity. No specific information about the relation between specific organic chemical compounds and the sensitivity under LC-HRMS conditions with electrospray ionisation was found. This result confirmed the perception that there is little information available about the relation between physicochemical properties and the MS sensitivity of a compound.

3 Analytical approach

3.1 Introduction

An LC-Orbitrap MS screening was performed for a reference set of 223 compounds by KWR and a LC-Q-ToF MS screening was performed to analyze 294 compounds by Vitens. The dataset is used to study the sensitivity of about 500 different individual compounds in more detail and to predict the response factor using chemical properties. 49 compounds were analyzed both by KWR and Vitens and are used to study the influence of the different Mass Spectrometers (Orbitrap from Thermo Scientific versus Q-ToF from ABSciex) and analysis conditions on the sensitivity of the individual compounds.

During the progress of the project the approach changed. First the approach was to classify a set of compounds into different chemical groups using relevant properties; analysis should be performed on a selection of compounds from each group. From the literature study more than ten properties were found to influence the response factor using electrospray ionisation (ESI). Creating classes of compounds with similar chemical properties for over 10 variables was however problematic. Therefore we changed the approach and decided not create classes beforehand. The approach proceeded with a large experimental dataset from LC-HRMS screening. The experimental data were classified into groups with different response factors. In a multivariate approach the classes with different response factors were then linked to all selected physicochemical properties.

3.2 Selection of compounds

The study makes use of two datasets:

1. 218 target compounds analysed with LC-Orbitrap MS at the KWR laboratory
2. 263 target compounds analysed with LC-Q-ToF MS at the Vitens laboratory

The list of 218 KWR target compounds is composed of 120 prioritized suspects (Sjerps et al. 2016), 97 compounds regularly found in surface waters (Jansen 2009) and one additional compound (pyrazole). The 120 compounds from the Sjerps study have been selected because these compounds have been prioritized from suspect screening LC-HRMS and need further confirmation of their identity with reference standards (Sjerps et al. 2016). The 97 other compounds have been selected because of their occurrence in surface water. These so called ‘top-100’ standard was analysed within each analytical LC-HRMS screening run at the KWR laboratory for the last 10 years for measurement of the recovery. Pyrazole is a recently identified substance with wide occurrence in Dutch surface waters and therefore a nice reference compound to test the applicability of the LC-HRMS screening. See Attachment II for the full list of KWR compounds.

The five internal standards atrazine-d5, bentazone-d6, chloroxuron, fenuron and neburon were added to each analytical run for quantification purposes.

The total list of KWR compounds covers a broad range of physicochemical properties; this is illustrated in Figure 3-1.

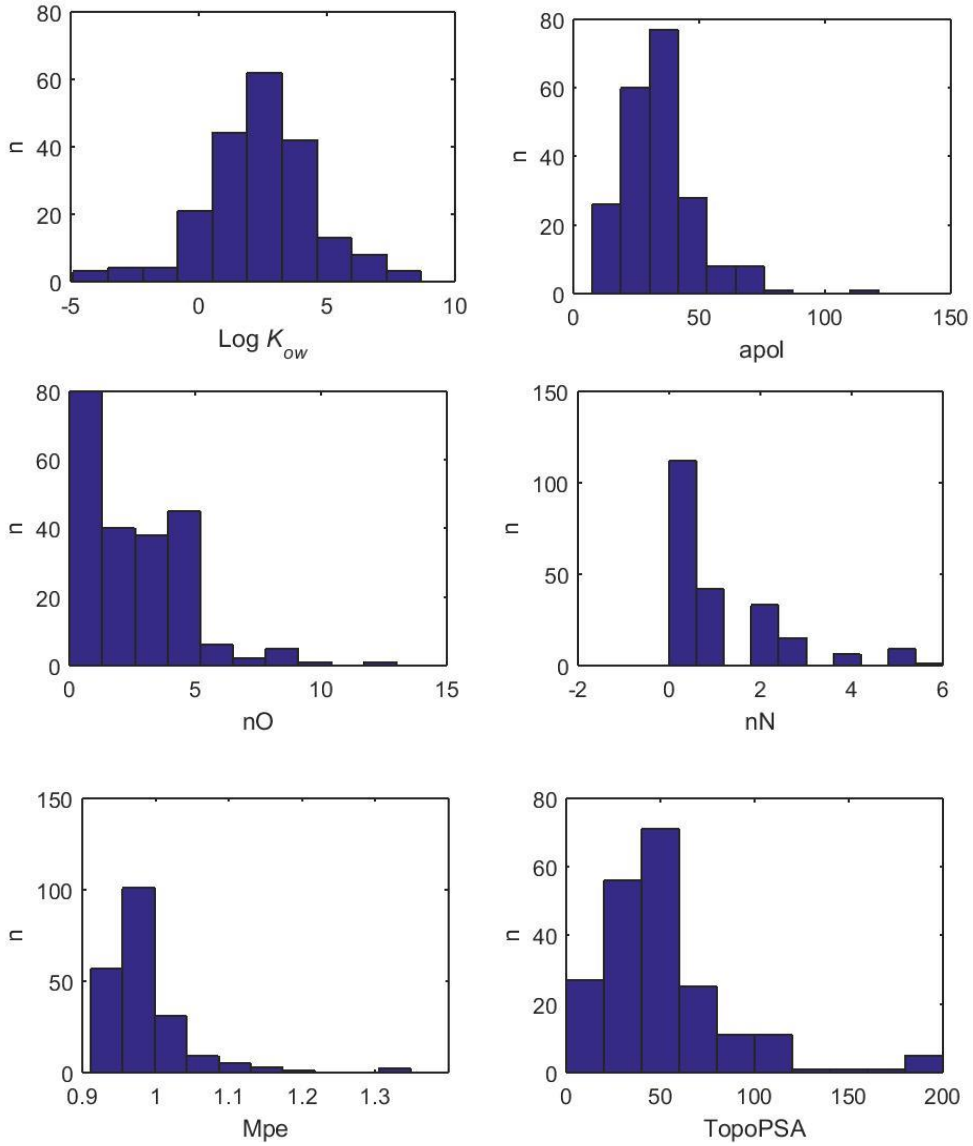


FIGURE 3-1 DISTRIBUTION OF A SELCTION OF SIX PHYSICOCHEMICAL PROPERTIES FOR THE COMPOUNDS ANALYSED BY KWR.

logKow= octanol-water distrubution coefficient, apol=measure for polarizability, nO=amount of oxygen atoms in molecule, nN=amount of nitrogen atoms in the molecule, Mpe=mean Pauling electronegativity, TopoPSA=topological polar surface area.

The list of 263 Vitens compounds is composed of a selection of target compounds included in the regular LC-HRMS screening (with a Triple Quad mass spectrometer)and based on the availability of physicochemical properties. See Attachment III for the full list of Vitens compounds.

3.3 Analysis with LC-HRMS

KWR

The individual stock compounds were dissolved in acetonitrile and diluted to a concentration between 3 en 15 mg/l in a mixture of 75% water and 25% acetonitrile.

The KWR compounds were analysed using direct injection of the standard, and analysed with conditions as described in Hogenboom et al (2009) and the LOA-600 protocol for LC-HRMS accurate mass screening with the Orbitrap (Thermo Scientific). The Orbitrap MS was used in the data dependent mode with a mass resolution of 7.500 and 30.000 with collision energy (CE) of 35% for both resolutions and CE of 65% specific for the resolution of 30.000. The obtained mass spectra of the 134 prioritized suspect compounds are stored in the Massbank database. In the BTO-report of the project 'Bevestigen Suspects', a protocol is included with a description of how to input the MS/MS spectra of chemical compounds into Massbank.

Injection of all individual compounds with and without an LC-column (Reversed Phase, Xbridge C18, particle size 3,5 µm, 2.1 x 150 mm) were performed using an injection loop of 10 µl. As internal standards to calculate a relative response, the compounds atrazine-d5, bentazon-d6 were added to all individual standards. The concentration of all internal standards was 1 mg/l. So the absolute amount of the individual compounds was 30-150 ng, the 1 absolute amount of the internal standards was 1 ng.

The following linear gradient was applied: starting at 5% acetonitrile/95% water with 0.1% formic acid (v/v/v) increasing to 99% acetonitrile/1% water with 0.1% formic acid in 40 min with a flow rate of 300 µl/min and a column temperature of 21˚C.

To study the influence of the chromatographic separation on the sensitivity of the individual compounds, all measurements are performed with direct injection and analysis with a LC-column.

Vitens

The individual stock compounds were solved in methanol or acetonitrile and diluted to a concentration between 0.5 to 1 µg/l in a solution consisting of water with acetonitrile.

The compounds were analysed by LC-Q-ToF MS on an AB Sciex instrument (API 5600). The conditions are in short: ion spray voltage: 4.5 kV, LC-column: Waters, XSELECT HSS C18 (Reversed Phase), 150 x 4.6 mm, particle size 3.5 µm, column temperature: 35˚C

The injection volume was 1000 µl, so the absolute amount of the individual compounds is about 0.5-1 ng.

The following linear gradient was applied: starting at 1% acetonitrile/99% water with 0.1% formic acid (v/v/v) increasing to 99% acetonitrile/1% water with 0.1% formic acid in min with a flow rate of 1000 µl/min.

As internal standards to calculate a relative response, the compounds atrazine-d5 (positive ionisation) and, bentazon-d6 (negative ionisation) were used.

3.4 Data handling

The internal standards atrazine-d5 and bentazon-d6 have been used for quantification purposes from the start in 2006 of the LC-Orbitrap MS screening at KWR. Concentrations of the detected compounds are expressed in µg/l internal standard equivalents. These two compounds were also used for the present study to calculate the relative response for each analyte.

The analyses with both direct injection and injection with the use of an LC-column are distributed across many measurement series that took place on different days, measured within a time period of about 2 months. Because the sensitivity of the mass spectrometer varies over time, the measured response expressed as peak area of the protonated ($[M+H]^+$) or deprotonated ($[M-H]^-$) molecules, is corrected for the response of the internal standard. This is performed by dividing the peakarea of the analyte by the peakarea of the internal standard atrazine-d5 (positive ions) or bentazone-d6 (negative ions). Adduct ions (e.g. sodium) are not registered.

The relation between the sensitivity for detection and the physicochemical properties was studied using LC-HRMS screening data from 6 subsets:

1. 218 compounds analysed at KWR with LC-column in positive ionisation mode
2. 218 compounds analysed at KWR with LC-column in negative ionisation mode
3. 218 compounds analysed at KWR without LC-column (direct injection) in positive ionisation mode
4. 218 compounds analysed at KWR without LC-column (direct injection) in negative ionisation mode
5. 263 compounds analysed at Vitens with LC-column in positive ionisation mode
6. 263 compounds analysed at Vitens with LC-column in negative ionisation mode

The sensitivity was linked to physicochemical properties from the following selection:

A. Physicochemical properties selected from literature and expert judgement (section 2.6).
B. Physicochemical properties described by 1300 descriptors calculated with the software PaDEL that were not selected from literature but show a high correlation with the response factor. The selected properties include:

1. Acidity/basicity (pKa/pKb, dependent on the positive or negative ionisation)*
2. Hydrophobicity (log P)
3. Molecular weight
4. Molecular size (molecular volume)
5. Number of hydrogen atoms that can attach to the surface area (polar surface area)
6. Number of N en O, S , P and Cl atoms atoms. The number of F, Br en I atoms are not taken into account.
7. Number of acid groups
8. Ionisation potential
9. Electronegativity
10. Number of double bond units (measure for aromaticity)
11. Dipole moment
12. Proton affinity

* No data could be retrieved for the pKa/pKb without additional costs. Therefore, this property was skipped in this study. It is recommended to include this parameter in a possible follow-up research.

The physicochemical properties were determined either by the software package PaDEL (Yap, 2011) or MOPAC (Stewart, 1990). For the software package PaDEL, a SDF file of a compound was used as input file. The SDF files were retrieved from the CACTUS (CADD Group Chemoinformatics Tools and User Services) web server. For the software package MOPAC, MOPAC2012 was used with the following options: PM6 NOMM STATIC mullik GEO-OK. Before MOPAC was executed, the SDF files were converted to MOL files using the openbabel software package (O'Boyle et al. 2011).

PaDEL *is a software package for calculating molecular descriptors and fingerprints. The software currently calculates 797 descriptors (663 1D, 2D descriptors, and 134 3D descriptors) and 10 types of fingerprints. These descriptors and fingerprints are calculated mainly using The Chemistry Development Kit. Some additional descriptors and fingerprints were added, which include McGowan volume, ring counts and count of chemical substructures.*

MOPAC *(Molecular Orbital PACkage) is a semiempirical quantum chemistry program based on Dewar and Thiel's NDDO approximation*

The relation between the physicochemical properties and the sensitivity was studied using the following tools:

- Boxplots distribution physicochemical properties for compounds that were detected and compounds that were not detected
- Spearman correlation coefficient
- Non-linear classification model
- Prediction of the detectability with the classification model

Using visual boxplots and a statistical test (t-test) it was evaluated whether individual physicochemical properties were different between compounds that could be detected and the compounds that could not.

The Spearman correlation coefficient was used to study the relationship between sensitivity and individual physicochemical properties from selection A (literature and expert judgement) and selection B (PaDEL). The Spearman rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function (linear or not). The sensitivity of detection was expressed as the response factor (peak area normalised for the internal standard and transformed to a10-log scale). A response factor larger than zero means that the sensitivity of the compounds is larger than the internal standard.

Next, the response factors of the analysed compounds were classified into two or four classes. For the two classes, the classification showed whether or not a signal will occur, whereas for the four classes, the classification also showed the magnitude of the response (Table 3-1). The combined set of physicochemical properties from selection A (literature and expert judgement) and selection B (PaDEL) was used to predict whether a compound was or was not detected with ESI (no response/response) and in which class the response factor of the chemical could be classified. The multivariate classification analysis is performed with a non-linear Gradient Boosting Classifier (Friedman, 2001) implemented in the scikit-learn machine learning toolbox of Python (Pedregosa et al., 2011). The performance is expressed

by the predictability score, determined by the performance of 50 iterations, in which for each iteration a randomly selected test set was chosen (25% of the original data set). The average R^2 of the test set for the 50 iterations was used as performance indicator. This express the fraction of well predicted classes.

TABLE 3-1 CLASSIFICATION OF CHEMICALS BASED ON EXPERIMENTALLY OBTAINED RESPONSE FACTORS.

Response class	Response factor	Pos direct injection	Neg direct injection	Pos with LC	Neg with LC
No response#	0	28%	69%	30%	70%
Small response factor	0-0.03 (<-1.5 log)	31%	17%	28%	16%
Medium response factor	0.03-1 (-1.5 log-0 log)	32%	15%	36%	14%
High response factor##	>1 (>0 log)	8%	2%	6%	3%

the response class 'no response' means: no response with the used instrument, analysis conditions en concentrations of the individual compounds. With the use of another instrument, different conditions or a higher concentration of the standard, a (small) response is possible.

the highest response factor is influenced by a lot of instrumental parameters and the used compounds. De highest response is bounded by e.g. the maximum ionisation rate and sensitivity in the mass spectrometer.

Next, the obtained classification model was applied to another list of 163 chemicals that are relevant for drinking water (Baken et al. 2015) to predict their detectability. For the complete list, see attachment IV.

4 Results and discussion

This chapter describes the performance of the LC-HRMS screening conducted by KWR and Vitens and whether the detectability could be predicted.

4.1 Performance of the LC-HRMS screening: internal standards

The analysis of the KWR target compounds was performed during a four month period. To correct for some drift of the analyte response in time, a correction for the response for an internal standard is applied.

4.1.1 Reproducibility of the peak area

The peakarea of the chromatographic peak is registrated to test the reproducibility of the three internal standards within the analytical series. The reproducibility of the peak area of the internal standards over a four month period (KWR data, subsets 1-2) is shown in Table 4-1

TABLE 4-1 REPRODUCIBILITY OF THE THREE INTERNAL STANDARDS IN LC-HRMS SCREENING. NEBURON IS USED AS AN ALTERNATIVE STANDARD FOR BOTH POSITIVE AND NEGATIVE IONISATION.

Ionisation mode	Standard	LC column	Average Peak area	STDEV	Relative stdev (%)	Number of detections
POS	atrazine-d5	no	1.40E+08	3.70E+07	26	212
		yes	2.34E+08	4.81E+07	21	206
	neburon	no	7.42E+07	1.96E+07	26	169
		yes	1.01E+08	5.97E+07	59	62[1]
NEG	bentazon-d6	no	5.14E+07	2.74E+07	53	210
		yes	5.47E+07	8.97E+06	16	212
	neburon	no	1.45E+07	4.10E+06	28	211
		yes	2.96E+07	6.62E+07	223[2]	212

[1] a selection of neburon measurements were used.
[2] this value seems unlikely. We could not found a good explanatory for this high value.

The relative standard deviation was satisfactorily low for atrazine-d5 (21-26%, n=206-212). With the use of the LC-column, bentazon-d6 measurements had a low relative standard deviation (16%), but without LC-column the relative standard deviation is high (53%). The relative standard deviation for neburon as an alternative compound for both positive and negative ionisation is relatively high, and varies from 26% (positive ionisation with direct injection) to 59% (positive ionisation with LC-column).

In Figure 4.1 the absolute peak areas of atrazine-d5 and bentazon-d6 for the analysis with and without LC-column are shown (KWR data, subsets 1-2)

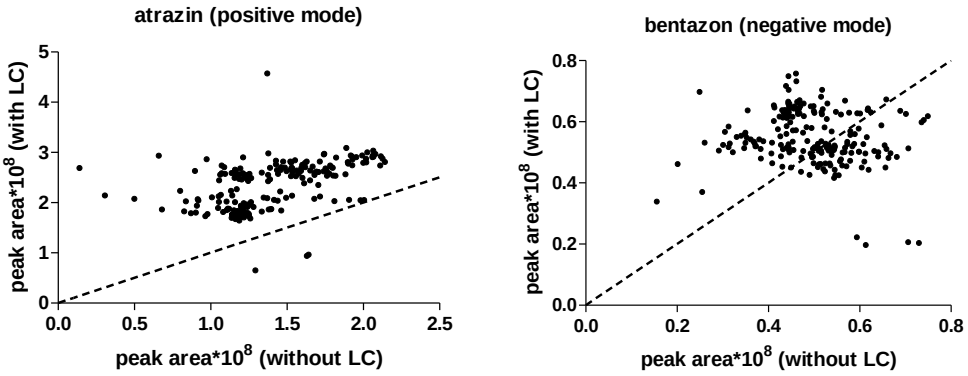


FIGURE 4-1 PEAKAREA OF ATRAZINE-D5 (LEFT) AND BENTAZON-D6 (RIGHT) IN THE LC-HRMS SCREENING WITH A LC COLUMN COMPARED TO DIRECT INJECTION AT IDENTICAL CONCENTRATION LEVEL. THE DOTTED LINE IS THE 1:1 LINE.

From this figure we can conclude that the averaged peak areas of atrazine-d5 with a LC-column were higher than peak areas with direct injection. A clear explanation of this difference in peak area could not be found. Maybe the percentage of water and the pH of the LC-eluens may play a role. For bentazon-d6, no clear differences between analysis with a LC-column and direct injection was observed.

The peakareas in both ionisation modes must be intense to use neburon as an alternative internal standard, for both positive and negative ionisation It was concluded from the experiments that the peakarea of neburon was in the same order as the internal standards atrazine-d5 and bentazon-d6 and therefore intense enough to use this compound as an alternative internal standard. The intensity of neburon in the positive ionisation mode was about 6 times higher compared to the negative ionisation mode. This is illustrated in Figure 4.2.

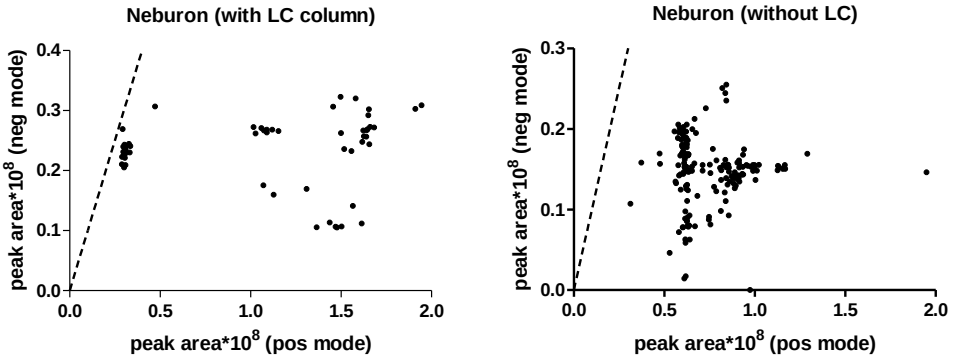


FIGURE 4-2 PEAK AREAS OF NEBURON IN THE POSITIVE AND NEGATIVE MODE WITH THE KWR LC-HRMS SCREENING WITH USE OF THE LC COLUMN (LEFT) AND DIRECT INJECTION. THE DOTTED LINE IS THE 1:1 LINE.

This results are confirmed by the Vitens data (subsets 5-6). The signal in both positive and negative ionisation is intense enough to use neburon as an internal standard. The ratio for the response positive/negative ionisation is 3:1. This difference in reponse ratio is affected by the instrument (Orbitap MS at KWR and Q-ToF MS at Vitens). See Figure 4-3.

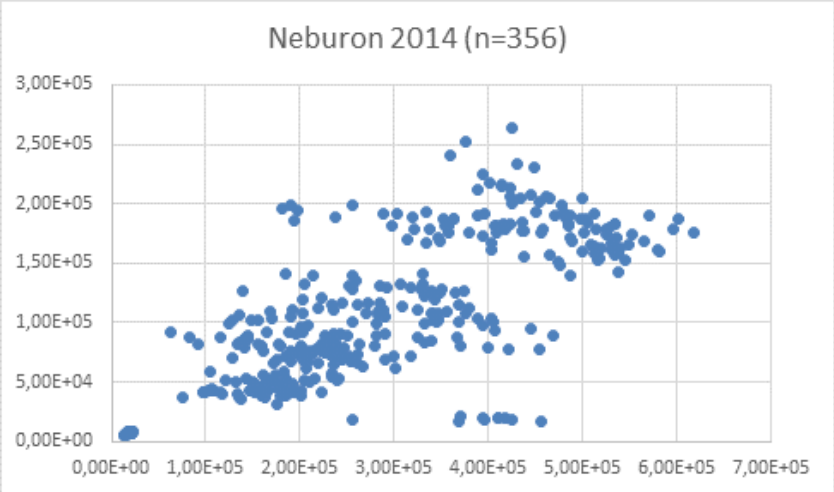


FIGURE 4-3 PEAK AREAS OF NEBURON IN THE POSITIVE (Y-AXIS) AND NEGATIVE (X-AXIS) MODE WITH THE LC-HRMS ANALYSIS AT VITENS

4.1.2 Reproducibility of the retention time

The retention time of the five internal standards atrazine-d5, bentazon-d6, fenuron, chloroxuron and neburon over the different analytical series is very stable (KWR data, subsets 1-2). The relative standard deviation is below 2% (see Table 4-2). De two labelled standards atrazine-d5 and bentazon-d6 were used to calculate the relative repons.

TABLE 4-2 REPRODUCIBILITY OF THE RETENTION TIME (

Ionisation mode	Standard	Retention time (min)	STDEV	Distribution (% stdev from average)	Number of detections
POS	atrazine-d5	15.8	0.2	1.2	199
	fenuron	8.7	0.2	1.9	192
	chlooroxuron	21.1	0.3	1.5	191
	neburon	24.5	0.0	0.1	13[1]
NEG	bentazon-d6	16.2	0.2	1.4	212
	chloroxuron	21.0	0.1	0.5	212
	neburon	24.0	0.1	0.3	212

[1] a selection of neburon measurements was used.

4.2 Performance of the LC-HRMS screening: studied compounds

Overall, more compounds could be detected with LC-HRMS screening coupled to ESI in the positive ionisation mode than in the negative ionisation mode (Table 4-3). In the positive ionisation mode, 67% of the compounds could be detected with LC-HRMS screening and 68% could be detected by direct injection. In the negative ionisation mode above 32% could be detected with LC-HRMS screening and 33% could be detected by direct injection.

TABLE 4-3 PERCENTAGE OF COMPOUNDS DETECTED IN THE POSITIVE AND NEGATIVE IONISATION MODE (KWR DATA, SUBSETS 1-2, N=218)

Ionisation mode	LC column	Not detected (%)	Detected (%)	Not analysed (%)
POS	no	31	68	1
	yes	33	67	1
NEG	no	65	33	1
	yes	67	32	1

The peak areas of the compounds were normalised relative to atrazine-d5 in the positive mode and bentazon-d6 in the negative mode. The response factors of almost 300 compounds varied between 10^{-6} and 10 times the internal standard (Figure 4-4-4). Less than 10% of the compounds had a response factor above 1; the ionisation efficiency of these compounds was larger than that of the internal standard. More than 90% of the compounds had a response factor below 1; this means that most of the concentrations expressed as the internal standards atrazine-d5 or bentazon-d6 were lower than the actual concentration.

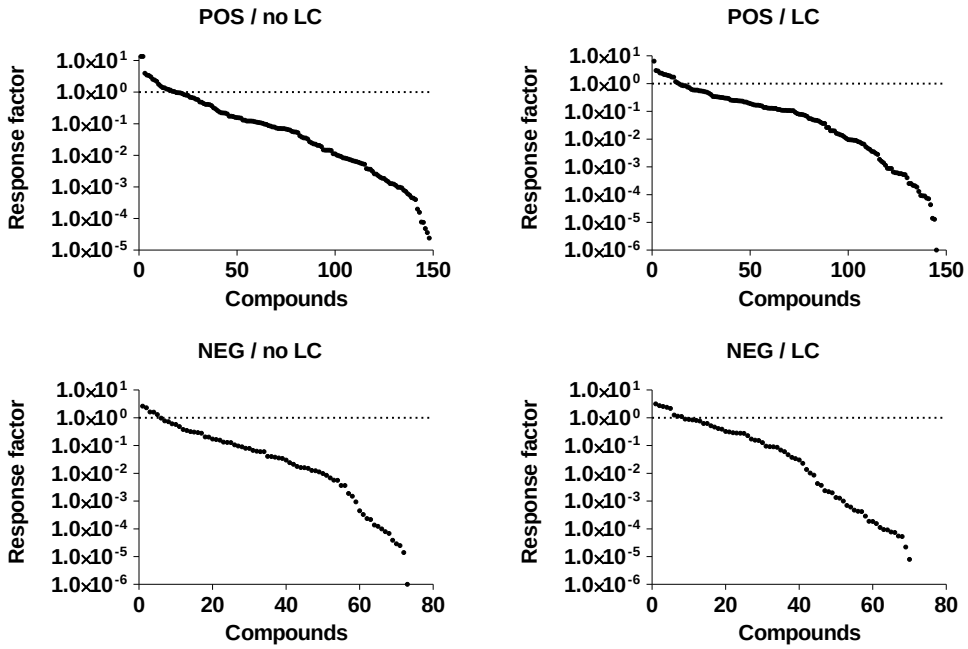


FIGURE 4-4 RESPONSE FACTORS OBTAINED WITH THE LC-HRMS SCREENING COUPELD TO ESI IN THE POSTIIVE (POS) AND NEGATIVE (NEG) IONISATION MODE, WITH DIRECT (NO LC) AND INDIRECT INJECTION (LC). (KWR DATA, SUBSETS 1-2)

43 compounds could not be detected at all (no signal in neither the positive nor negative ionisation mode).

17 compounds (see Table 4-4 for the list) show a (relative low) response with direct injection, while no response is observed with analysis using a LC-column, probably due to the high polarity of the compounds and as a result no retention on the LC-column. An example of this phenomena is the compound pyrazole (logKow = 0.26).

TABLE 4-4 COMPOUNDS ONLY DETECTED WITH DIRECT INJECTION (NOT WITH ANALYSIS USING A LC-COLUMN (KWR DATA, SUBSETS 1-4)

Compound	Response factor (relative to internal standard)	Log K_{ow}
Positive ionisation mode		
Aminomethylphosphonic acid [AMPA]	0.001849	-2.47
Nitrilotriacetic acid [NTA]	0.003790	-3.81
Pyrazole	0.121753	0.26
Decamethylcyclopentasiloxane	0.001222	8.03
6-Aminopenicillanic acid	0.002245	0.6
2,4-dichlooraniline	0.007847	2.78
Iopamidol	0.001262	-2.42
Iohexol	0.001109	-3.05
Triethoxyvinylsilane	0.000024	1.16
2,5-Dimethyl-2,5-di(tert-butylperoxy)hexane	0.000036	6.55
Negative ionisation mode		
Amidotrizoic acid	0.000099	1.37
Salicylic acid-2-ethyl-1-hexyl ester	0.000014	5.97
Nitrilotriacetic acid [NTA]	0.016048	-3.81
2,4-dichlooraniline	0.000937	2.78
Iopamidol	0.016100	-2.42
Iohexol	0.001898	-3.05
Monochloroacetic acid	0.003632	0.22

The comparison of the normalised peak areas (response factor) of the compounds in the screening with direct (without LC-column) and indirect injection (with LC column) is shown in Figure 4-5.

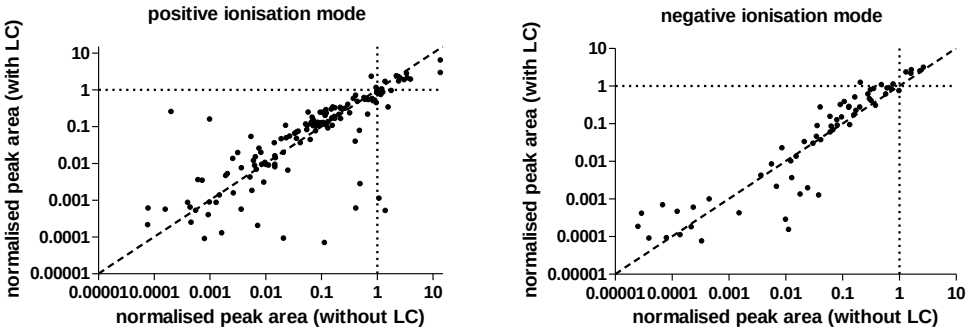


FIGURE 4-5 NORMALISED PEAK AREA (LOG SCALE) OF ALL COMPOUNDS IN THE LC-HRMS SCREENING (KWR DATA, SUBSETS 1-4))WITH THE USE OF A LC-COLUMN COMPARED TO DIRECT INJECTION.
LEFT: POSITIVE IONISATION, RIGHT: NEGATIVE IONISATION.
THE DOTTED LINE IS THE 1:1 LINE.

Since most compounds approached the 1:1 line, the difference in normalised peak area between direct injection and analysis with the use of a LC-column was similar, both for positive and negative ionisation. The largest part of the compounds ionised in the positive ionisation mode approach the 1:1 line; therefore most compounds had a larger peak area in the analyses with the use of a LC-column (similar to the internal standard atrazine-d5, Figure 4-1). In the negative ionisation mode, the studied compounds showed similar peak areas with direct and indirect injection (similar to bentazon-d6, Figure 4 1).

4.3 Performance of the LC-HRMS screening: Vitens and KWR compared

The two datasets of Vitens and the four datasets of KWR (subsets 1-6) contain 52 identical compounds, see for more information Attachment III. In the positive ionisation mode, the Vitens response factors (obtained at a Q-TOF instrument) were generally bigger than the KWR response factors (Orbitrap instrument). Due to the limited amount of data, no clear conclusions can be drawn for the negative ionisation mode.

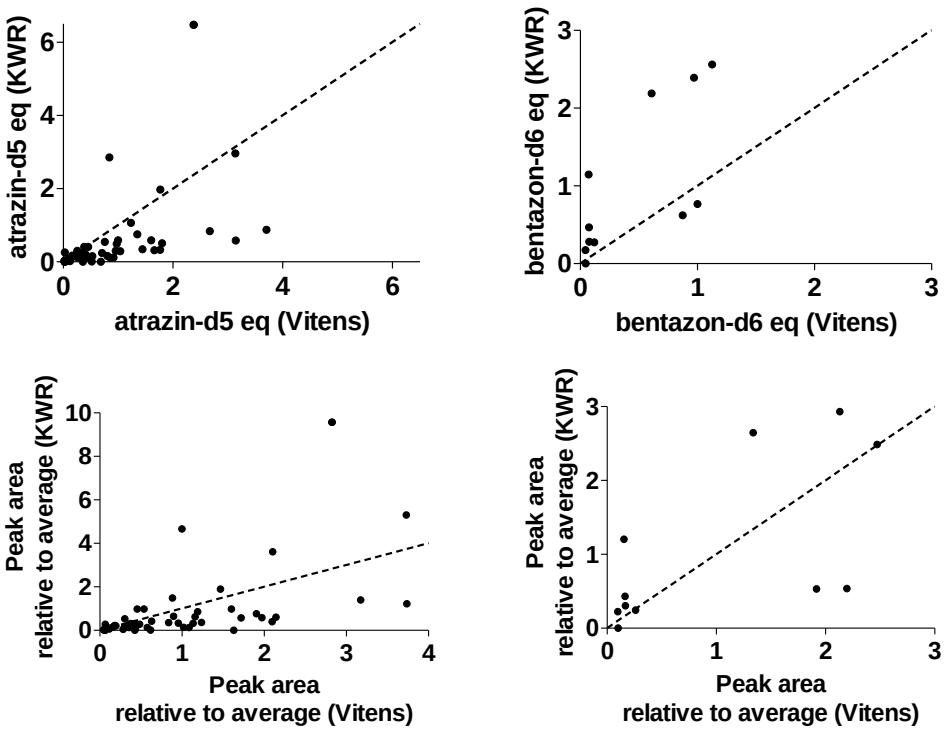


FIGURE 4-6 COMPARISON OF RESPONSE FACTORS OF DATA OBTAINED BY KWR AND VITENS, RESPECTIVELY. PEAK AREAS: RELATIVE TO INTERNAL STANDARD (ABOVE) OR AVERAGE PEAK AREA (BELOW) IN THE POSTIVE (LEFT) AND NEGATIVE (RIGHT) IONISATION MODE.

4.4 Correlation between physicochemical properties and response

The 12 physicochemical properties from A (literature and expert judgement, see paragraph 3.4) in combination with the 44 physicochemical properties from selection B (best correlating from all PaDEL descriptors, see Table 4-5) were related to the response factors. Beforehand, the 12 physicochemical properties from A were tested for correlation. Molecular weight (no.3) and the number of double bond units as a measure for aromaticity (no. 10) show no correlation and were skipped.

The 44 descriptors in Table 4-5 were tested for mutual correlations (orthogonality). After the removal of strongly mutually correlated descriptors (see Attachment V), 32 descriptors were used for further analysis (see 'final selection' column in Table 4-5).

In total 10 + 32 = 41 descriptors were used for the final correlation study between physicochemical properties and response factor.

From the calculated 1300 descriptors provided by the software program PaDEL, two descriptors that were not selected based on the literature survey, show a high correlation with the response: AATSC0e and GATS3e. Both descriptors are a measure for the electronegativity of the molecule.

TABLE 4-5 41 SELECTED PHYSICOCHEMICAL DESCRIPTORS WITH DESCRIPTION AND SOURCE.

Descriptors	Description	Source	Final selection
Log_Kow	n-Octanol/water partition coefficient – measure of Hydrophobicity	Episuite	Yes Log Kow
AlogP	Ghose-Crippen LogKow	PaDEL	yes
Henry	Volatility (Henrys law constant in atm-m3/mol)	Episuite	yes
McGowan_Volume	McGowan characteristic volume	PaDEL	yes
VABC	Van der Waals volume	PaDEL	no, cross correlated with McGowan_Volume
TopoPSA	Topological polar surface area	PaDEL	yes
DIPOLE	Dipole moment	MOPAC	yes
nHBAcc	Number of hydrogen bond acceptors	PaDEL	yes
nHBAcc2	Number of hydrogen bond acceptors	PaDEL	no, cross correlated with HBAcc
nHBAcc3	Number of hydrogen bond acceptors	PaDEL	no, cross correlated with HBAcc
nHBAcc_Lipinski	Number of hydrogen bond acceptors	PaDEL	no, cross correlated with HBAcc
nHBDon	Number of hydrogen bond donors	PaDEL	yes
nHBDon_Lipinski	Number of hydrogen bond donors	PaDEL	no, cross correlated with HBDon
nAcid	Number of acidic groups.	PaDEL	yes
nAtom	Number of atoms	PaDEL	no, cross correlated
nHeavyAtom	Number of heavy atoms (i.e. not hydrogen)	PaDEL	no, cross correlated
nC	Number of carbon atoms	PaDEL	yes
nH	Number of hydrogen atoms	PaDEL	yes
nN	Number of nitrogen atoms	PaDEL	yes
nO	Number of oxygen atoms	PaDEL	yes
nS	Number of sulphur atoms	PaDEL	yes
nP	Number of phosphorus atoms	PaDEL	yes
nF	Number of fluorine atoms	PaDEL	yes
nCl	Number of chlorine atoms	PaDEL	yes
nBr	Number of bromine atoms	PaDEL	yes
nI	Number of iodine atoms	PaDEL	yes
nX	Number of halogen atoms (F, Cl, Br, I, At, Uus)	PaDEL	yes
TopoSPA	Topological polar surface area	PaDEL	yes
apol	Sum of the atomic polarizabilities (including implicit hydrogens)	PaDEL	no, cross correlated
Mp	Mean atomic polarizabilities (scaled on carbon atom)	PaDEL	yes
Sp	Sum of atomic polarizabilities (scaled on carbon atom)	PaDEL	no, cross correlated with Mp
MLFER_S	Combined dipolarity/polarizability	PaDEL	yes

Descriptors	Description	Source	Final selection
IONISATION POTENTIAL	Ionisation potential	MOPAC	yes
Mi	Mean first first ionisation potentials (scaled on carbon atom)	PaDEL	yes
Si	Sum of first first ionisation potentials (scaled on carbon atom)	PaDEL	yes
Mpe	Mean atomic Pauling electronegativities (scaled on carbon atom)	PaDEL	yes
Spe	Sum of atomic Pauling electronegativities (scaled on carbon atom)	PaDEL	no, cross correlated with Mpe
HEAT OF FORMATION	Standard enthalpy of formation or the enthalpy change to form a mole of compound at 25^{o}C from its elements in their standard state	MOPAC	yes
TOTAL ENERGY	Sum of electronic and nuclear-nuclear repulsion energies for molecules, isolated in vacuum, without vibration at 0 K	MOPAC	yes
DBE	Double Bond Equavalents	Xcalibur	no, no correlation with response
MW	Molecular Weight	Xcalibur	no, cross correlated with number of different atoms (e.g. C, H, N,O)
AATSC0e	Average centered Broto-Moreau autocorrelation - lag 0 / weighted by Sanderson electronegativities	PaDEL	yes
GATS3e	Geary autocorrelation - lag 3 / weighted by Sanderson electronegativities	PaDEL	yes

First, to find out the distinctive character of the different descriptor values, KWR compounds with and without a response were used. The results of this automated study for compounds that ionise in the positive mode are shown in Figure 4-77. The results for compounds that ionise in the negative mode are shown and 4-7.

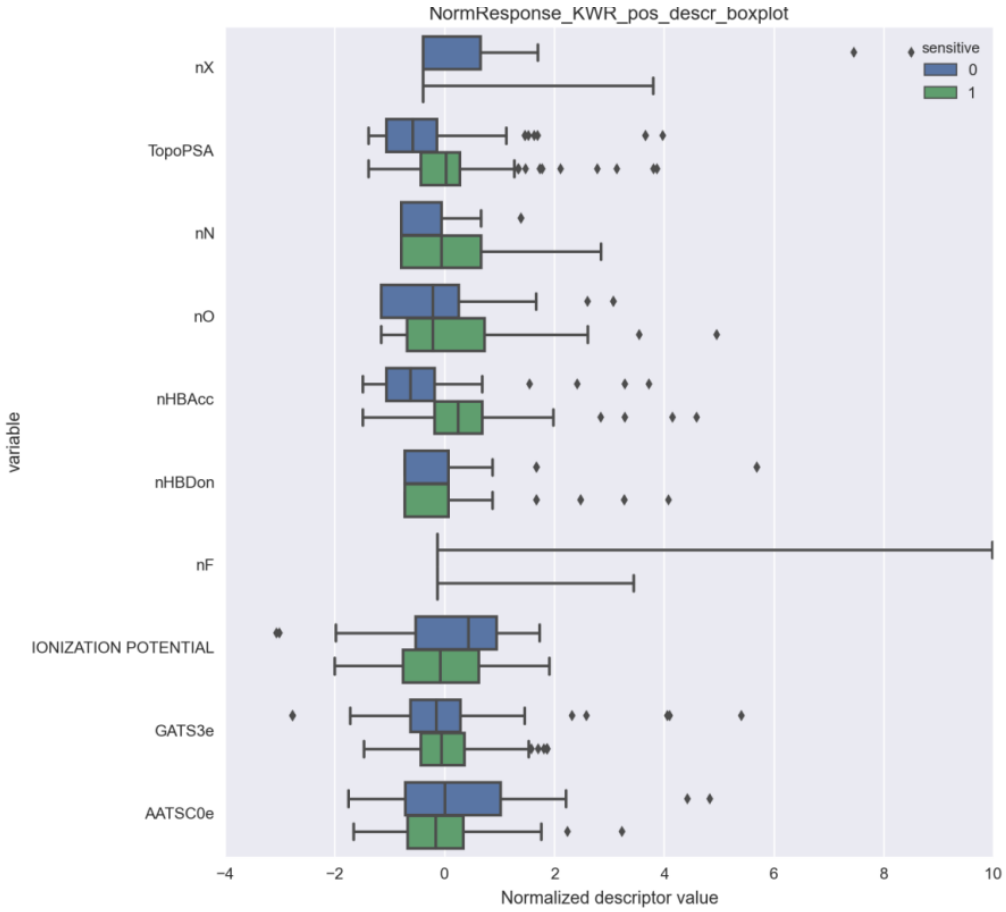


FIGURE 4-7 DISTRIBUTION (MEAN AND 90% PERCENTILE) OF SOME OF THE BEST CORRELATING DESCRIPTOR VALUES OF THE COMPOUNDS WITH A RESPONS (1: GREEN) AND WITHOUT A RESPONSE (0:BLUE) IN THE POSITIVE IONISATION MODE (KWR DATASETS).

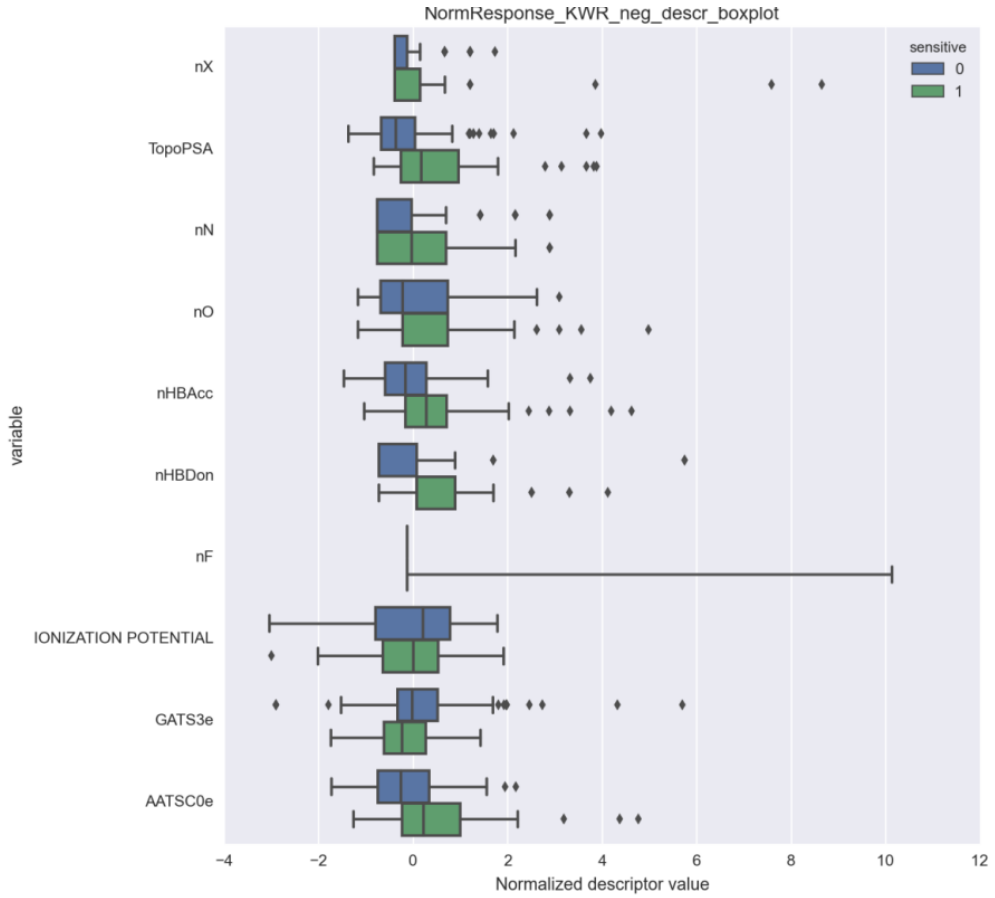


FIGURE 4-8 DISTRIBUTION (MEAN AND 90% PERCENTILE) OF THE SOME OF THE BEST CORRELATING DESCRIPTOR VALUES OF COMPOUNDS WITH A RESPONS (1: GREEN) AND WITHOUT A RESPONSE (0:BLUE) IN THE NEGATIVE IONISATION MODE (KWR DATASETS).

Most descriptor values show a large overlap in the boxplots and do not significantly differ for compounds with and without a response. The most distinctive descriptor value for positive ionisation is nHBacc, a measure for the number of Hydrogen Bond acceptors. The most distinctive descriptor value for negative ionisation is nHBDon, a measure for the number of Hydrogen Bond Donors.

Probably the effect of one single descriptor value cannot be fully distinctive, since the response of a specific compound depends on multiple descriptor values (physicochemical parameters).

Second, the 32 descriptors were correlated to the response factors in the six datasets (excluding the compounds without a response) as paragraph 3.4 (data handling).

Figure 4-99 and Attachment VI show the correlations between the physicochemical descriptors and the LC-HRMS response (compounds with no response excluded). A correlation of 1 is a total positive linear correlation (reddish colours), a correlation of 0 is no linear correlation, whereas a correlation of -1 is a total negative linear correlation (bluish colours). The chemical descriptors on the right hand side of the figure show the largest extent of correlation with the response. The highest correlations observed are in the range of 0.4-0.5 or -0.4--0.5, indicating that a single descriptor may explain at most about 40-50% of the variability of the normalized response. So there is no single descriptor that may

explain the variability in the normalized response to a large extent. The descriptors that show the best correlation with the response are highlighted in Figure 4-10 and Figure 4-3.

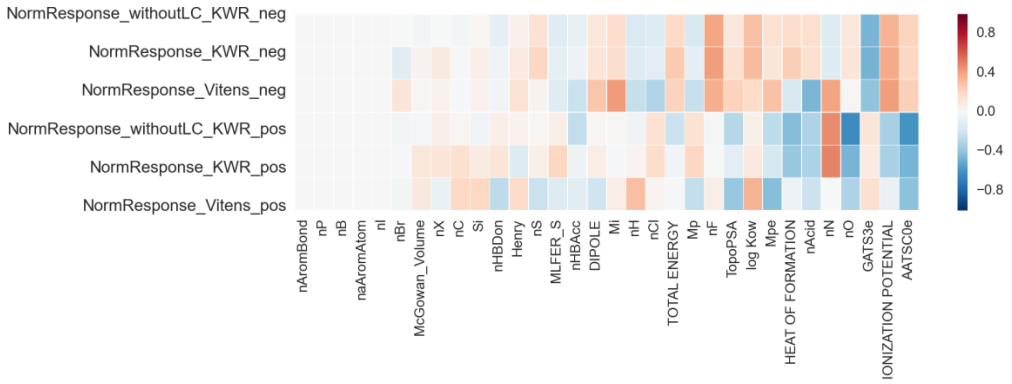


FIGURE 4-9 EXTENT OF CORRELATION (EXPRESSED BY THE DETERMINATION COEFFICIENT R^2) BETWEEN THE SELECTED PHYSICOCHEMICAL DESCRIPTORS AND THE RESPONSE FACTOR (LEFT TO RIGHT; LOW TO HIGH CORRELATION).

Average centered Broto-Moreau autocorrelation (AATSCOe, R^2=-0.6), the number of oxygen atoms (nO, R^2=-0.63) and the number of nitrogen atoms (nN, R^2=0.5) show the highest correlation with the response factor in LC-HRMS screening at KWR in the positive mode. The number of oxygen atoms and the number of nitrogen atoms show a negative and a positive correlation, respectively. This observation confirmed our experience that nitrogen containing compounds give a better response due to the higher proton affinity caused by the nitrogen atoms in the molecule that facilitates proton addition to form a protonated molecule ($[M+H]^+$). Oxygen in the form of a hydroxyl or ketone group has the opposite effect due to an increased electronegativity of the molecule.

In the negative ionisation mode, ionisation potential (R^2=0.38), the number of fluorine atoms (nF, R^2=0.42) and Geary autocorrelation (GATS3e, R^2=0.47) show a high correlation. The variation of the response was large for molecules that contain no fluorine atoms. However if a molecule contains one or more fluorine atoms the response was high due to the electronegativity of fluorine atoms and which facilitates hydride abstraction to form $[M-H]^-$ ions

Topological polar surface area (TopoPSA, R^2=-0.38), average centered Broto-Moreau autocorrelation (AATSCOe, R^2=-0.41) and the mean atomic Pauling electronegativity (MPe, R^2=-0.42) show a high correlation with responses obtained in LC-HRMS screening at Vitens in the positive ionisation mode. The response decreases when the topological polar surface area decreases, due to an increased electronegativity of the molecule.

In the negative ionisation mode, ionisation potential (R^2=0.42), mean first ionisation potentials (Mi, R^2=0.42) and the number of acidic groups (nAcid, R^2=-0.46) show a high correlation with observed response. As expected, the response increases when the ionisation potential increases.

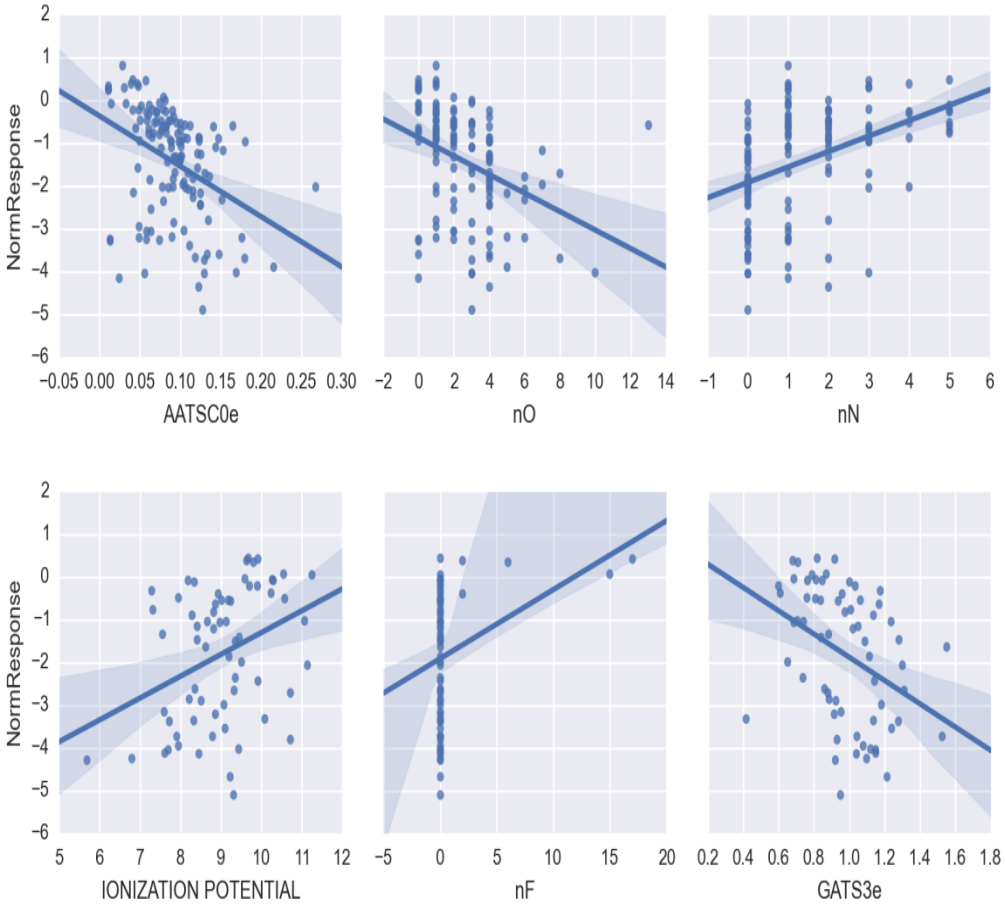


FIGURE 4-10 RELATION BETWEEN THE BEST CORRELATING PSYSICOCHEMICAL DESCRIPTORS AND THE RESPONSE FACTOR (LOG SCALE) IN THE POSITIVE (ABOVE) AND NEGATIVE IONISATION MODE (BELOW) IN THE SCREENING PERFORMED AT KWR. NOTE THAT A NEGATIVE NORMRESPONSE MEANS A RESPONSE LOWER THAN THE STANDARD (AS A RESULT OF THE LOG TRANSFORMATION SCALE)

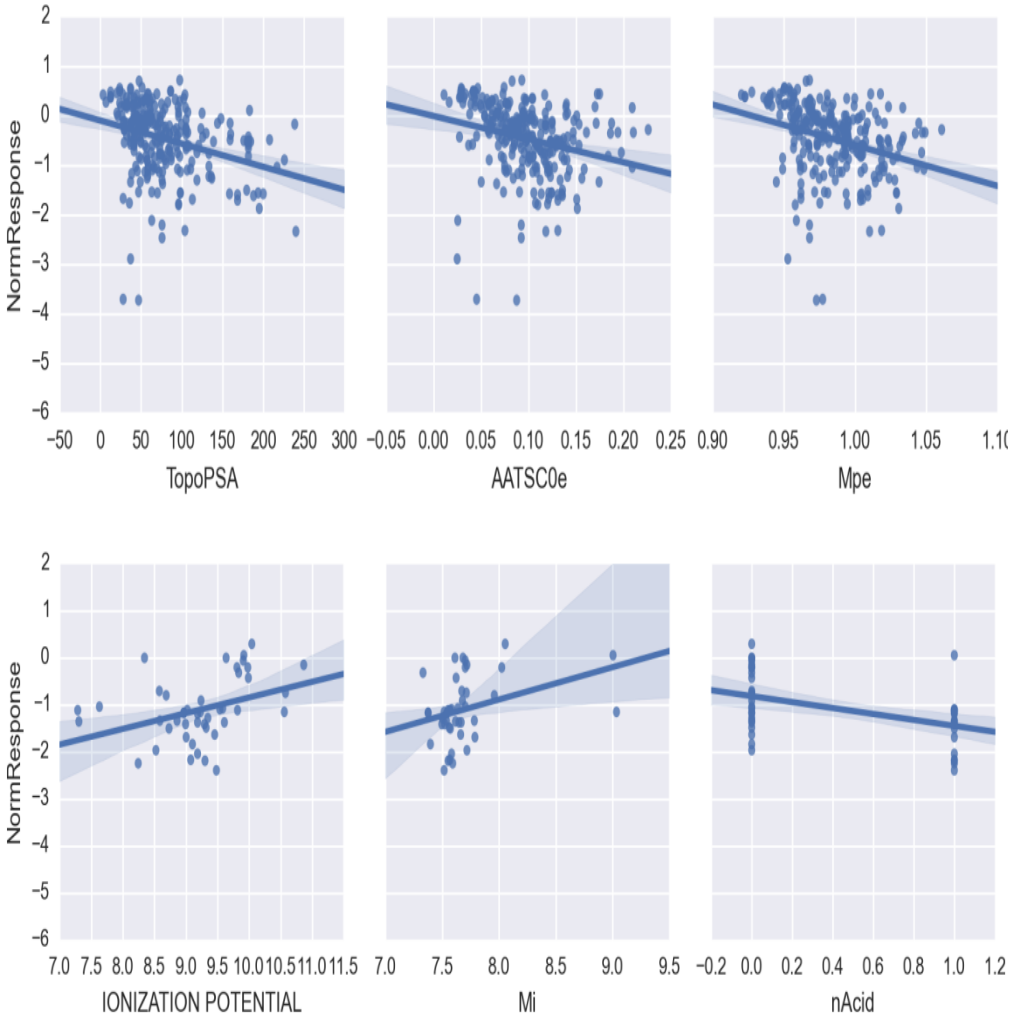


FIGURE 4-3 RELATION BETWEEN THE BEST CORRELATING PHYSICOCHEMICAL DESCRIPTORS AND THE RESPONSE FACTOR (LOG SCALE) IN THE POSITIVE (ABOVE) AND NEGATIVE IONISATION MODE (BELOW) IN THE SCREENING PERFORMED AT VITENS.

4.4.1 Examples

In the following section we highlight some examples comparing the response factor between related compounds. Note that this information is based on only one measurement. For more detailed conclusions, additional measurements are necessary.

Atrazine and transformation products

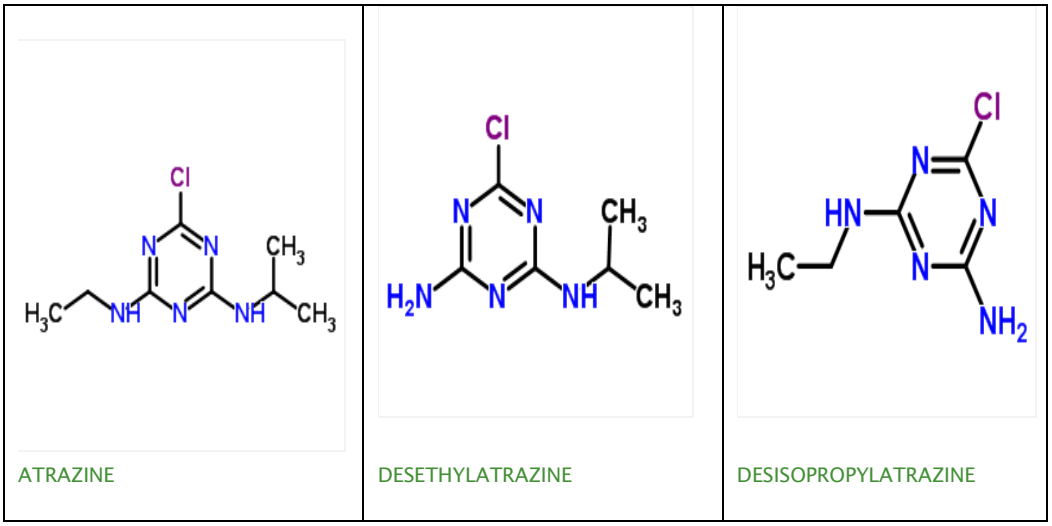


FIGURE 4-4 CHEMICAL STRUCTURE OF ATRAZINE AND TRANSFORMATION PRODUCTS.

Atrazine shows the largest response in the positive ionisation mode, up to 80% of atrazine-d5 (Figure 4-13). Both desethylatrazine as well as desisopropylatrazine have a smaller response; related to the smaller molecular weight because of the loss of one ethyl-group (C2H5) and one extra methyl-ethyl-group (C3H7), see Figure 4-12. As expected, the compounds were not ionised in the negative ionisation mode.

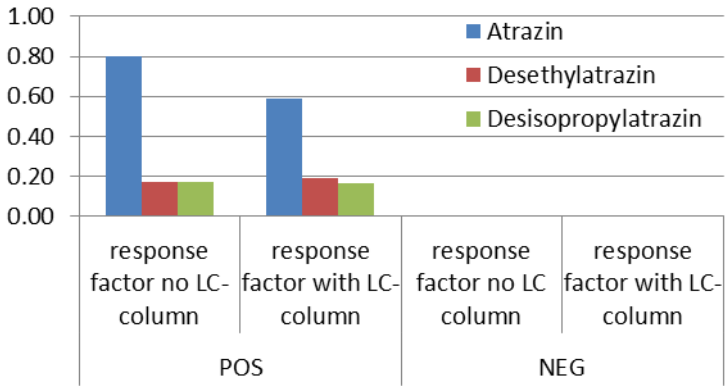


FIGURE 4-5 RESPONSE FACTOR OF ATRAZINE AND DERIVATIVES

Diuron and transformation products

Next, the herbicide diuron and its transformation products 1-(3,4-dichlorophenyl)-3-methylurea and 1-(3,4-dichloorfenyl)urea were considered, the chemical structures are presented in Figure 4-14. Diuron shows the largest response in the positive ionisation mode, up to 20% of atrazine-d5 (Figure 4-). Both 1-(3,4-dichlorophenyl)-3-methylurea and 1-(3,4-dichloorfenyl)urea have a smaller response; related to the smaller molecular weight because of the loss of one or two methyl-groups (CH_3). The response in the negative ionisation mode shows an inversed pattern (when analysed with LC-column).

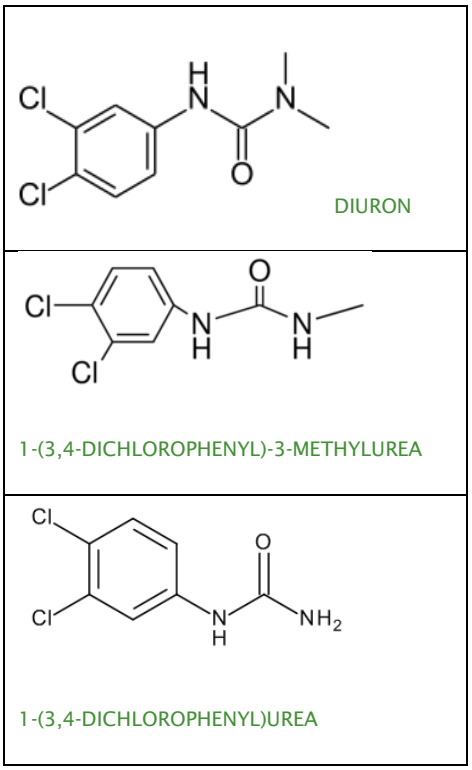


FIGURE 4-6 CHEMICAL STRUCTURESSTRUCTURE OF DIURON AND TRANSFORMATION PRODUCTS

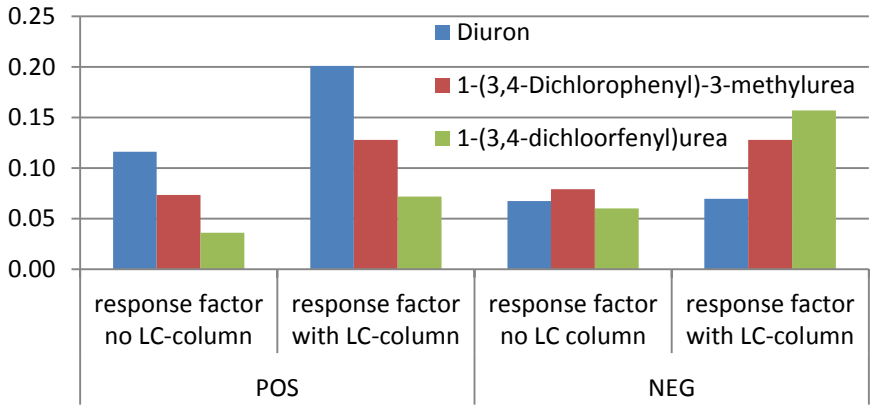


FIGURE 4-15 RESPONSE FACTOR OF DIURON AND SOME TRANSFORMATION PRODUCTS

Glymes

The solvents tetraglyme, triglyme and diglyme are polyoxyethyleneglycol dimethyl ethers (glymes) of different chain lengths (Figure 4-7). We cannot observe a clear pattern of response factors for these compounds (Figure 4-87). Following direct injection the response factor of the three compounds was equal: 0.01-0.02% of that of atrazine. Following injection by LC-HRMS, the response factor of triglyme exceeds that of the two others. No response was observed for any of the glymes in the negative ionisation mode.

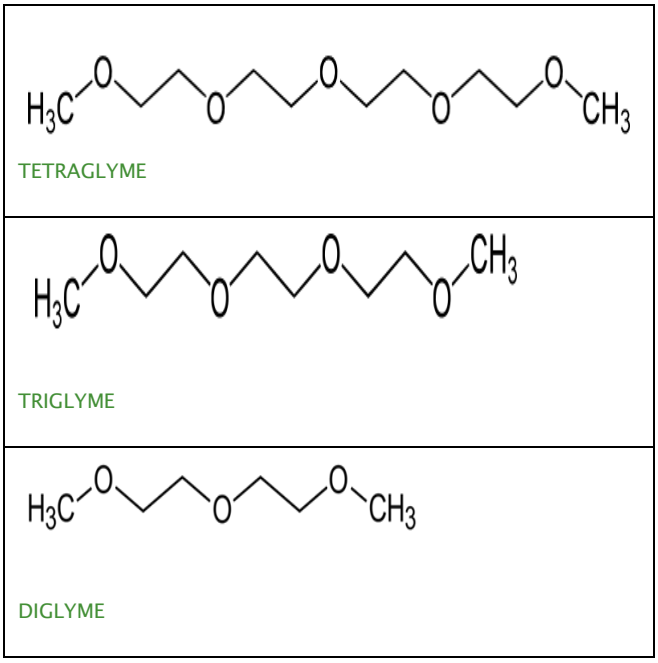


FIGURE 4-7 CHEMICAL STRUCTURE OF 3 POLYOXYETHYLENEGLYCOL DIMETHYL ETHERS

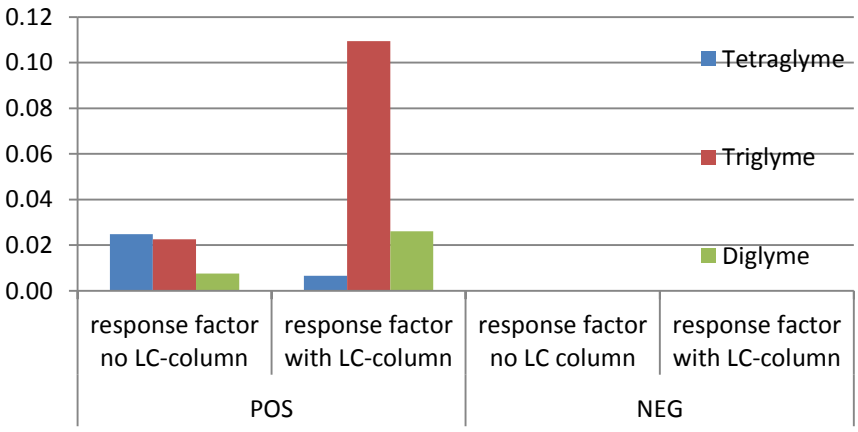


FIGURE 4-8 RESPONSE FACTOR OF 3 POLYOXYETHYLENEGLYCOL DIMETHYL ETHERS (GLYMES).

Alkylphosphates

Triethylphosphate and tributylphosphate are esters of phosphoric acid, for the chemical structures see Figure 4-18.

The compound with the longest alkylchains, tributylphosphate had the highest response (Figure 4-10). Triethylphosphate had a smaller response, related to the smaller molecular weight because of the shorter chains. No response was obtained in the negative mode.

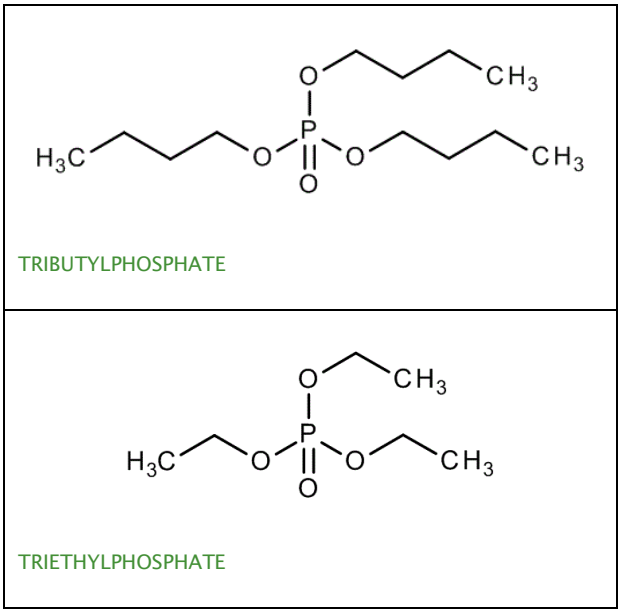


FIGURE 4-98 CHEMICAL STRUCTURE OF TRIBUTYLPHOSPHATE (ABOVE) AND TRIETHYLPHOSPHATE (BELOW).

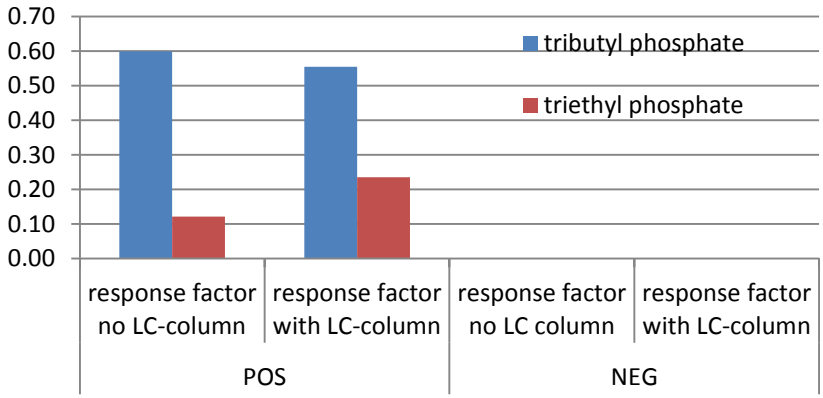


FIGURE 4-109 RESPONSE FACTOR OF TWO ALKYLPHOSPHATES

4.4.2 Relation between LC-retention time and hydrophobicity

As expected from literature (Barron et al. 2016, Bade et al. 2015), the retention time shows an excellent relation with the hydrophobicity (expressed as log K_{ow} (see Fig. 4-20 the right dark red colums; KWR: R^2 =0.82, Vitens: R^2=0.77). The more hydrophobic the substance (the larger the log K_{ow} value), the more the substance will be retained by the LC-column and the longer the retention time.

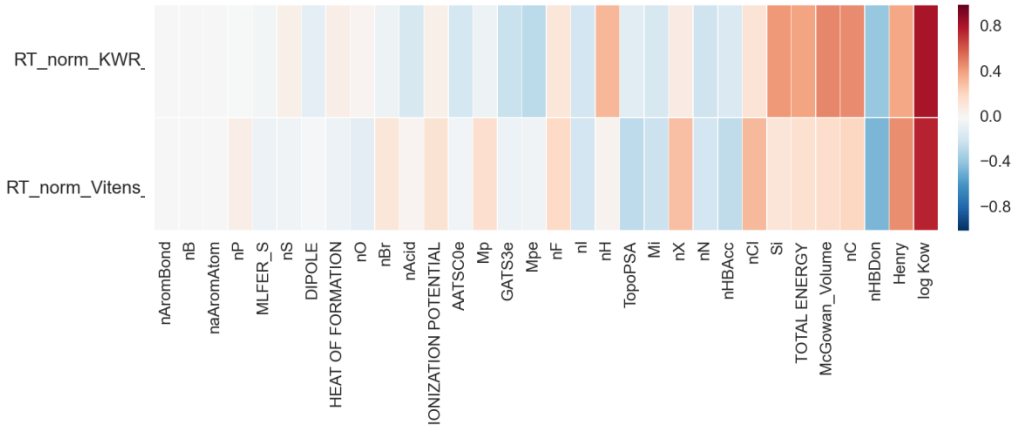


FIGURE 4-20 EXTENT OF CORRELATION (EXPRESED BY THE DETERMINATION COEFFICIENT R^2) BETWEEN THE SELECTED PHYSICOCHEMICAL DESCRIPTORS VALUES AND THE RETENTION TIME (LEFT TO RIGHT; LOW TO HIGH CORRELATION COEFFICIENT).

For all the Vitens compounds, the logKow is plotted against the retention time.

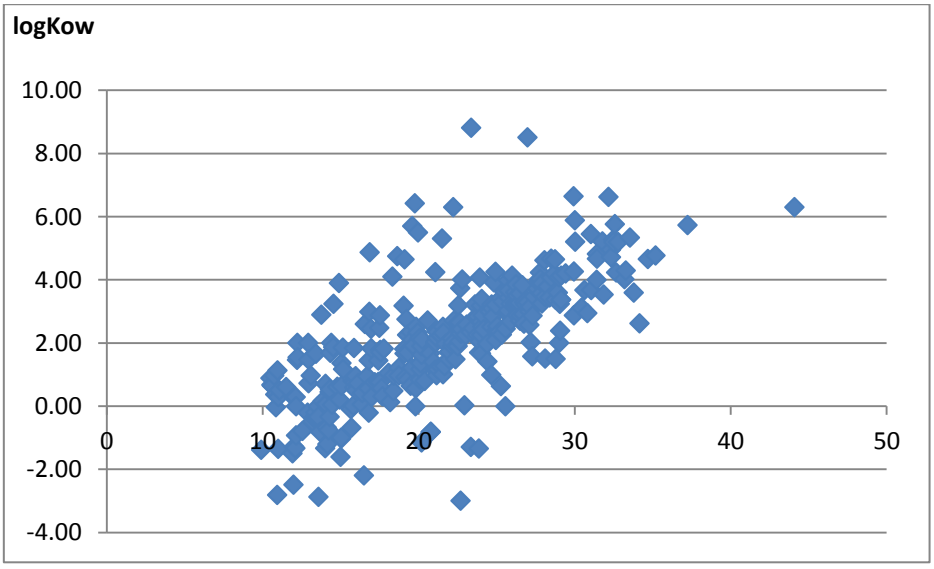


FIGURE 4-11 CORRELATION BETWEEN LOGKOW AND RETENTION TIME FOR THE VITENS COMPOUNDS.

Although less significant, we found an additional good relation with the Henry constant (KWR: R^2 =0.38, Vitens: R^2=0.47) as a measure for volatility both for the Vitens data and the

KWR data. When the Henry coefficient of a substance is high, the substance is more retained on the LC-column.

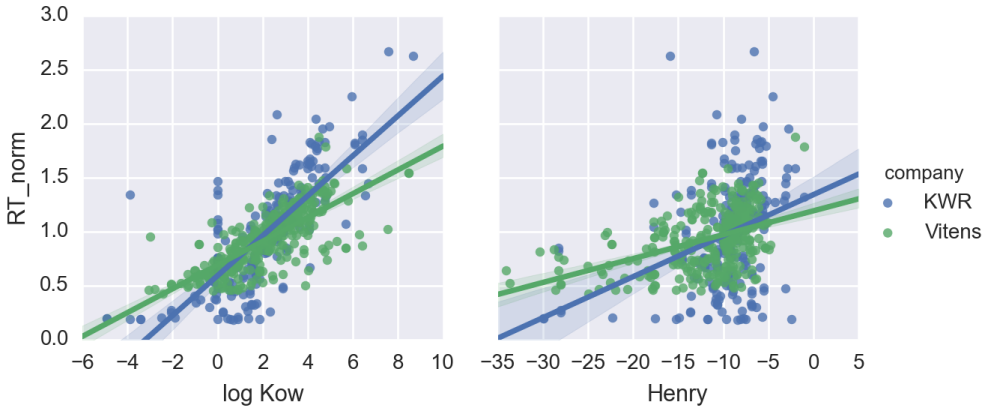


FIGURE 4-22 RELATION BETWEEN (I) HYDROPHOBICITY (LOG KOW) AND (II) VOLATILITY (HENRY CONSTANT) AND THE RETENTION TIME DETECTED IN THE LC-HRMS SCREENING BY KWR AND VITENS BOTH IN THE POSITIVE (GREEN) AND NEGATIVE (BLUE) IONISATION MODE

4.5 Classification model

To investigate their correlation with response, 32 physicochemical properties were selected. The selection included physicochemical properties from literature and expert judgement (selection A) and the parameters from PaDEL (selection B) that show a good correlation. These properties were used to describe whether a compound did or did not show a response with ESI (no response/response) and in which class the response factor of the chemical could be classified.

The classification analysis for the KWR LC-HRMS screening data was performed for two classification schemes:

- Two classes: compounds with (i) no response and compounds (ii) with a response.
- Four classes: compounds with (i) no response, (ii) low response, (iii) medium response and (iv) high response (see table 4-6)

First, the 200 compounds were classified into two classes: compounds that show a response and compounds without a response (Table 4-6). With Python a non-linear classification model was applied to predict the substance class using the 32 physicochemical descriptors.

Python (programming language)

Python is a widely used high-level, general-purpose, interpreted, dynamic programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale. Python is managed by the non-profit Python Software Foundation.

Three quarters of the experimental data were used to train the classification model. The predictability score was determined using a randomly selected test set including 25% of the original data set. The 50 iterations resulted in a predictability score of 79% to 81%; this

implies that whether a compound can or cannot be detected was predicted properly for 80% of the compounds. In general a R^2 of 0.7 or higher for a test set is regarded as a good performance of a statistical model (Wols et al., 2012). Note that for a classification of two classes a prediction score of 50% is equal to a random guess for response/no response, so that the prediction score of 80% is the minimum value for a good prediction model.

TABLE 4-6 TWO CLASSES OF CHEMICALS BASED ON EXPERIMENTAL OBTAINED RESPONSE FACTORS.

Compounds	Pos no LC	Neg no LC	Pos with LC	Neg with LC
Total data set	194	200	194	200
Test set	49	50	49	50
No response (RF=0)	55	133	58	135
With response (RF>0)	139	67	136	65
R^2	0.79	0.83	0.81	0.79
R^2 std	0.05	0.05	0.06	0.04

The contribution of each descriptor to the model is shown in Attachment VII (Contribution of descriptors values in the prediction of the response factor). The contribution is expressed as the average and standard deviation of the contribution by predicting the response factor class in 50 repetitions. Most contributing descriptors were volatility (Henry), ionisation potential (Si), hydrophobicity (log Kow or AlogP), hydrogen bond donors (HBDon), number of acidic groups (nAcid) and the combined dipolarity/polarizabilit (Milfer_S).

Second, the compounds were classified into four classes of different response factors relative to internal standard: no response, low response, average response, high response (Table 4-7). Again, a non-linear classification model was applied to predict the substance class using the 32 physicochemical descriptors. The predictability score became 60% to 76% depending on the ionisation mode and LC column. The scores were lower than the predictability score for the two classes (80%). Here, the prediction score of 25% is equal to a random guess for a response class. The borders chosen for the small, reduced or large response are a bit arbitrary, which may explain why the model shows a reduced performance. Since the probability to predict a wrong response factor is thus 25-40%, still some improvements are required in this model to use it to predict the response class for other chemicals. Alternatively, a (non-linear) regression model could be used to directly predict the response factor instead of using classes.

TABLE 4-7 FOUR CLASSES OF CHEMICALS BASED ON EXPERIMENTALLY OBTAINED RESPONSE FACTORS.

Response class	Response factor	Pos no LC	Neg no LC	Pos with LC	Neg with LC
Total data set	-	194	200	194	200
Test set	-	49	50	49	50
No response	0 (no response)	55	133	58	135
Small response factor	0-0.03 (<-1.5 log)	60	33	55	31
Medium response factor	0.03-1 (<-1.5log-o log)	63	30	70	28
High response factor	>1 (>0 log)	16	4	11	6
R^2		0.62	0.76	0.60	0.69
R^2 std		0.07	0.05	0.06	0.06

The contribution of each descriptor to this model is shown in Attachment VI. Most contributing descriptors were again volatility (Henry), hydrophobicity (log Kow or AlogP), hydrogen bond donors (HBDon) and the combined dipolarity/polarizability (Milfer_S). Other contribution descriptors are the average centered Broto-Moreau autocorrelation (AATSCo) and heat of formation.

4.6 Prediction of detectability of drinking water relevant compounds

Using the prediction model from section 4.5, a prediction was made for the LC-HRMS detectability of a total list of 163 drinking water relevant compounds (Baken et al. 2015). The selection was based on compounds that were detected in drinking water. From these 163 compounds, physicochemical descriptor values (necessary for the prediction of detectability) were available for 150 out of 163 compounds.

The prediction (no response/response) for the 150 compounds was again performed in 50 iterations (similar as for the training of classification model described in section 4.5). From these 50 iterations, the probability to be detected is expressed as a percentage. For a probability of 20% or less it is assumed the compound has no response. A probability of over 80% indicates that the compounds have a response. The compounds for which the prediction varies over time (a probability between 20-80%) are not reliable.

In Figure 4-23is shown that 55% of the compounds can be detected in the positive mode (with a probability of more than 80%) and 25% in the negative mode (with a probability of more than 80%). In total, 65% of the 150 drinking water relevant compounds were predicted to be sensitive for detection with LC-HRMS screening coupled to ESI.

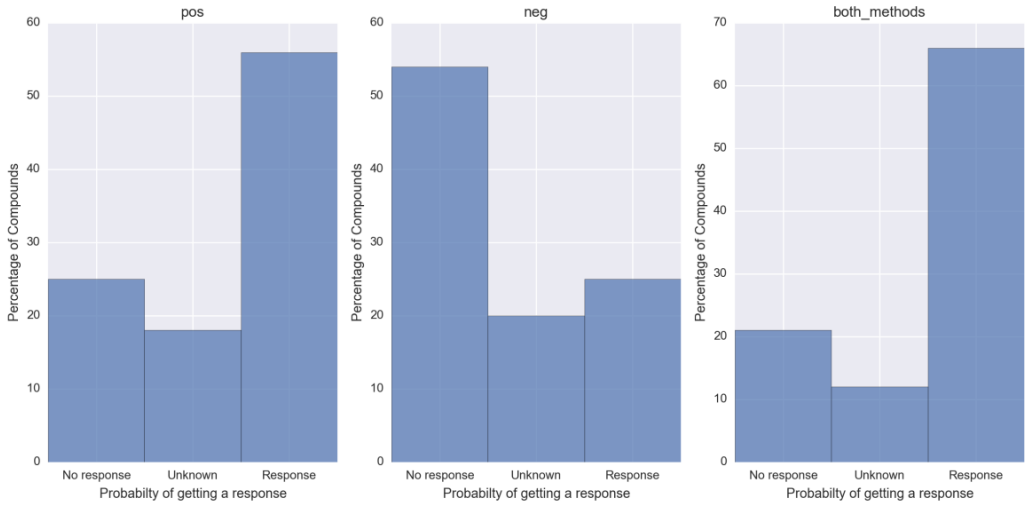


FIGURE 4-23 COMPOUND DISTRIBUTION OF THE PREDICTABILITY TO GET A RESPONSE. NO RESPONS (<20%), UNKNOW (20-80%), OR RESPONSE >80% TO BE DETECTED IN THE POSITIVE AND NEGATIVE IONISATION MODES AND WITH BOTH MODES.

As an example, the sensitivity of pyrazole was compared to all other compounds. Pyrazole occurred in the summer of 2015 in the river Meuse at high concentrations.

The response factor of pyrazole is 12% (compared to atrazine-d5) and therefore this compound is detectable with LC-HRMS by direct injection. 25% of the compounds (54 individuals) have a larger response factor; and 75% a smaller response factor. Pyrazole could

not be detected with analysis using a LC-column as used in this study because of the high polarity (see Table 4-4).

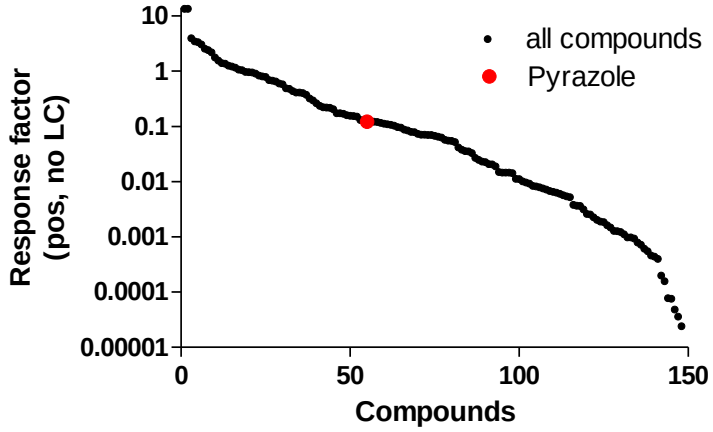


FIGURE 4-24 RESPONSE FACTOR OF THE COMPOUNDS DETECTED WITH POSITIVE IONISATION MODE AND DIRECT INJECTION; RELATIVE TO PYRAZOLE.

5 Conclusions and recommendations

5.1 Conclusion

The LC-HRMS analysis of almost 500 compounds showed that the response factor varied between 10^{-6} and 10 times that of the internal standard. More than 90% of the compounds had a response factor below 1, this means that the response is lower than the internal standard. Concentration expressed as equivalents internal standard (atrazine-d5 for positive ionisation or bentazon-d6 for negative ionisation) were generally an underestimation of the actual concentration.

Correlation analysis showed that the retention time is strongly related to log K_{ow} en the Henry constant. In the positive ionisation mode the response factors were significantly related to the number of oxygen and nitrogen atoms in the analyte, topological polar surface area and electronegativity. The number of fluorine atoms, the ionisation potential and the number of acidic groups were all significantly correlated to the response factor in the negative ionisation mode.

The non-linear classification model using the 32 physicochemical descriptors was able to predict with 80% accuracy whether a compound could be detected by LC-HRMS or not. The most contributing descriptors were Henry coefficient, ionisation potential (Si), hydrophobicity (log Kow or AlogP), hydrogen bond donors (HBDon), number of acidic groups (nAcid) and the combined dipolarity/polarizability (Milfer_S).

Whether or not a compound belonged to a specific response factor class (no response, low, reduced and high response) could be predicted with an accuracy of 60%-76%. Since the probability to predict a wrong response factor is thus 25-40%, still some improvements are required in this model to use it to predict the response class for other chemicals.

The classification model was applied to 150 drinking water relevant compounds. 65% of these compounds were predicted to be sensitive for detection with LC-HRMS screening coupled to ESI. Confirmation of the compounds, predicted to be detectable by experiments, was not performed in this study but seems reliable based on expert knowledge.

5.2 Recommendations

We have shown that a classification model can be used for the prediction of the response factor of specific compounds analyzed by LC-HRMS with an ESI-interface. However, refinement of the model could increase the certainty of the correct prediction.

Some recommendations for refinement of the model are:

- Collection of more measurement data;
- Inclusion of missing physicochemical descriptors (such as pKa/pKb) in the model;
- Gathering more insight in the influence of matrix effects (e.g. caused by humic acids);
- Gathering more insight in the influence of the pH during ionisation;
- Gathering more insight whether water facilitate the ionisation process;
- Gathering more insight in the role of specific chemical groups present in the molecule (e.g. amine, hydroxyl):
- Next to classification of four classes, a regression analysis could be performed to predict the response factor for new chemicals.

6 Future monitoring and modeling

Organic micropollutants are still a major concern for drinking water companies, recently again underlined by the high concentrations of pyrazole observed in the river Meuse water. Monitoring of drinking water sources is optimal with a complete chemical screening. This screening includes non-target screening for the detection of known and unknown compounds and effect directed screening with the use of in vitro bioassays.

Modelling can improve our knowledge about the behaviour and effects of substances. In this study we developed a tool to predict the sensitivity of individual chemical compounds when analysed by non-target LC-ESI-MS screening, based on physicochemical descriptors. Besides the detection of the compounds, prediction models for the toxicity of the detected compounds and for the removal efficiency of different treatment processes to produce drinking water are important to estimate the risk of chemical compounds for drinking water production.

The ultimate goal of these tools is an optimal safeguarding of the water quality of drinking water and their sources. Combining the predictive models for screening, toxicity and water treatment will be a powerful tool for water companies to directly assess the occurrence, toxicity and removal of new organic micropollutants (Baken et al. 2014 and 2015).

7 References

Bade R, Bijlsma L, Sancho JV, Hernández F. Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water. Talanta 2015;139:143-149.

Bade R, Bijlsma L, Miller TH, Barron LP, Sancho JV, Hernández F. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis. Sci Total Environ 2015;538:934-941.

Baken, K., Kools, S.,(2014). Innovative testing strategies and their relevance for evaluating chemical drinking water quality. BTO 2014.009

Baken, K., Schriks. M, Sjerps., R.(2015). Toxicologische risicobeoordeling geprioriteerde stoffen, BTO 2015.056

Barron LP, McEneff GL. Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods. Talanta 2016;147:261-270.

Chalcraft, K.R., Lee, R., Mills, C. and Britz-McKibbin, P. (2009) Virtual quantification of metabolites by capillary electrophoresis- electrospray ionisation-mass spectrometry: Predicting ionisation efficiency without chemical standards. Analytical Chemistry 81(7), 2506-2515.

Cech, N.B. and Enke, C.G. (2000) Relating electrospray ionisation response to nonpolar character of small peptides. Analytical Chemistry 72(13), 2717-2723.

Chemical Abstract Service, http://www.cas.org

Dole M, Mack LL, Hines RL, Mobley RC, Ferguson LD, Alice MB (1968). "Molecular Beams of Macroions". Journal of Chemical Physics. 49 (5): 2240–2249.

Birendra N. Pramanik; A.K. Ganguly; Michael L. Gross (28 February 2002). Applied Electrospray Mass Spectrometry: Practical Spectroscopy Series. CRC Press. pp. 4–. ISBN 978-0-8247-4419-9.

Ehrmann, B.M., Henriksen, T. and Cech, N.B. (2008) Relative Importance of Basicity in the Gas Phase and in Solution for Determining Selectivity in Electrospray Ionisation Mass Spectrometry. Journal of the American Society for Mass Spectrometry 19(5), 719-728.

European Chemicals Agency Information on Chemicals
http://echa.europa.eu/web/guest/regulations/reach/substance-registration/publishing-information-from-dossiers

Espinosa, M.S., Folguera, L., Magallanes, J.F. and Babay, P.A. (2015) Exploring analyte response in an ESI-MS system with different chemometric tools. Chemometrics and Intelligent Laboratory Systems 146, 120-127.

Friedman, J. (2001) Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol. 29, No. 5, 2001.

Ghosh B, Jones AD. (2015). Dependence of negative-mode electrospray ionisation response factors on mobile phase composition and molecular structure for newly-authenticated neutral acylsucrose metabolites. Analyst, 140, 6522-31

Hogan Jr, CJ, Fernández De La Mora, J. (2009). Tandem ion mobility-mass spectrometry (IMS-MS) study of ion evaporation from ionic liquid-acetonitrile nanodrops. Physical Chemistry Chemical Physics, vol. 11, issue 36, 8079-8090.

Hogenboom AC, van Leerdam JA, de Voogt P. Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography-hybrid linear ion trap Orbitrap mass spectrometry. J Chromatogr A 2009;1216(3):510-519.

Henriksen, T., Juhler, R.K., Svensmark, B. and Cech, N.B. (2005) The relative influences of acidity and polarity on responsiveness of small organic molecules to analysis with negative ion electrospray ionisation mass spectrometry (ESI-MS). Journal of the American Society for Mass Spectrometry 16(4), 446-455.

Hellmuth C, Weber M, Koletzko B, Peissner W. (2012). Nonesterified fatty acid determination for functional lipidomics: comprehensive ultrahigh performance liquid chromatography-tandem mass spectrometry quantitation, qualification, and parameter prediction. Anal Chem 84, 1483-90

Huijzer, J. , (2006). samenstellen top 100 prioritaire stoffen in rivierwater, opbouwen LC-DAD database, KWR afstudeerverslag 2006.

Jansen, R., (2009) Inventarisatie van frequent aanwezige stoffen in oppervlaktewater en optimalisatie van het KWR screening protocol, KWR afstudeerverslag 2009.

Kruve, A., Kaupmees, K., Liigand, J. and Leito, I. (2014) Negative electrospray ionisation via deprotonation: Predicting the ionisation efficiency. Analytical Chemistry 86(10), 4822-4830.

Van Leerdam, J.A., Puijker. L.M., Hogenboom, A.C., Oomen, C.A.H.M., Gruter, H.F.M., van Genuchten C.C.A. (2008). Onderzoek naar organisc he microverontreinigingen in winningen van waterbedrijf Oasen. KWR rapport 08.013

ter Laak, T.L., Puijker, L.M., van Leerdam, J.A., Raat, K.J., Kolkman, A., de Voogt, P. and van Wezel, A.P. (2012) Broad target chemical screening approach used as tool for rapid assessment of groundwater quality. Science of the Total Environment 427-428, 308-313.

Leito, I., Herodes, K., Huopolainen, M., Virro, K., Künnapas, A., Kruve, A. and Tanner, R. (2008) Towards the electrospray ionisation mass spectrometry ionisation efficiency scale of organic compounds. Rapid Communications in Mass Spectrometry 22(3), 379-384.

Nguyen, S. and Fenn, J.B. (2007) Gas-phase ions of solute species from charged droplets of solutions. Proceedings of the National Academy of Sciences of the United States of America 104(4), 1111-1117.

O'Boyle, N. M., M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison (2011). "Open Babel: An open chemical toolbox." Journal of Cheminformatics 3(1): 33.

Oss, M., Kruve, A., Herodes, K. and Leito, I. (2010) Electrospray ionisation efficiency scale of organic compound. Analytical Chemistry 82(7), 2865-2872.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot

and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research 12: 2825-2830.

Reemtsma T, Berger U, Arp HPH, Gallard H, Knepper TP, Neumann M, et al. Mind the Gap: Persistent and Mobile Organic Compounds - Water Contaminants That Slip Through. Environ Sci Technol 2016;50(19):10308-10315.

Yap, C. W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011 May;32(7):1466-74.

Schwarzenbach, R.P.; Escher, B.I.; Fenner, K.; Hofstetter, T.B. Johnson, C.A.; von Gunten, U.; Wehrli, B. (2006). The challenge of micropollutants in aquatic systems. Science, 313, (5790), 1072-1077

Speksnijder, P., Kolkman, A. (2012), Toepassingen van hydrophilic intercation chromatografie (HILIC) voor de analyse van zeer polaire stoffen in water. BTO 2012.020. Nieuwegein. KWR.

Stewart, J. J. P. (1990). "MOPAC: A semiempirical molecular orbital program." Journal of Computer-Aided Molecular Design 4(1): 1-103

Vughs, D. Kooij, P. Kolkman, A. (2015), HILIC sreening – analyse van zeer polaire stoffen in water. BTO 2015.076. Nieuwegein. KWR.

Wols, B. A., D. Vries and C. H. M. Hofman-Caris (2012). Removal efficiency of pharmaceuticals in drinking water production: application of QSARs. BTO 2012.228(s). Nieuwegein, KWR.

Attachment I : recovery of top-100 compounds

Recovery of top-100 compounds (Jansen, 2009). Isolation of the compounds with Oasis HLB. Elution with 100% acetonitrile

Compound	SPE Recovery (n=2)
anhydro-erythromycin-A (metabolite of erythromycine)	139 ± 1
carbamazepin	131 ± 5
metoprolol	125 ± 5
trifenylfosfineoxide (=TPPO)	125 ± 5
erythromycine A	121 ± 1
pentoxifylline	118 ± 8
chloortoluron	112 ± 2
diuron	111 ± 3
isoproturon	109 ± 9
metazachloor	109 ± 11
dimethomorf	106 ± 6
terbutylazin	102 ± 3
fenazon	101 ± 0
pirimicarb	100 ± 0
dichlorprop (2,4-DP)	98 ± 0
linuron	97 ± 5
metribuzin	97 ± 5
2,4-dichloorfenoxyazijnzuur (=2,4-D)	96 ± 1
desisopropylatrazine	95 ± 1
metobromuron	95 ± 1
mecoprop [=MCPP]	93 ± 10
1-(3,4-dichloorfenyl)-3-methylureum	91 ± 6
1-(3,4-dichloorfenyl)ureum (metabolite of diuron)	91 ± 6
atrazine	91 ± 3
monuron	89 ± 8
desethylatrazine	88 ± 9
simazin	87 ± 4
tetraethyleenglycol dimethyl ether (=tetraglyme)	85 ± 28
caffeine	84 ± 2
bezafibraat	83 ± 8
metoxuron	78 ± 0
chlorpyrifos	77 ± 0
bromacil	75 ± 1
sulfadimidine	75 ± 5
ethofumesate	74 ± 10

Compound	SPE Recovery (n=2)
sulfamethoxazool	73 ± 6
metolachloor	72 ± 18
tri-n-butylfosfaat	70 ± 19
2,4-dichlooraniline	69 ± 7
4-chloor-2-methylfenoxyazijnzuur (=MCPA)	68 ± 6
chloridazon	68 ± 2
tris-(2-chloorpropyl)fosfaat	66 ± 9
2,6-dichloorbenzamide (=BAM)	65 ± 3
diclofenac	64 ± 10
diisobutylftalaat	61 ± 4
iopromide	56 ± 2
2,4-dichloorfenol	55 ± 5
anhydroerythromycin-B (metabolite erythromycine)	53 ± 3
2-(1,1-dimethylethyl)-4,6-dinitrofenol (=dinoterb)	50 ± 8
2-methyl-4,6-dinitrofenol (=DNOC)	50 ± 7
bentazon	50 ± 8
carbendazim	46 ± 1
2,4,6-trichloorfenol	43 ± 10
azinfos methyl	40 ± 5
dichlobenil	38 ± 17
2-aminoacetofenon	37 ± 9
triethylfosfaat	37 ± 19
2,4-dinitrofenol	36 ± 8
triethylglycol dimethyl ether (=triglyme)	31 ± 3
dicamba	13 ± 1
iomeprol	10 ± 1
iopamidol	10 ± 1
4,4'-sulfonyldifenol (=bisfenol-S)	0
bisfenol-A	0
diethyleentriaminepentaazijnzuur (=DTPA)	0
diethylene glycol dimethyl ether (=diglyme)	0
naftaleen-1-sulfonaat	0
parathion-methyl	0
sotalol	0
tetrachloor orthoftaalzuur	0

Attachment II: KWR list of studied target compounds

CAS number	Compound name	Origin
101-42-8	Fenuron	Internal Standard
555-37-3	Neburon	Internal Standard
1982-47-4	Chlooroxuron	Internal Standard
163165-75-1	Atrazine-d5	Internal Standard
25057-89-0	Bentazon-d6	Internal Standard
1002-69-3	chloordecaan	Top 100
100-42-5	Ethenylbenzene [Styrene]	Top 100
1007-28-9	Desisopropylatrazine	Top 100
101-83-7	dicyclohexylamine	Suspect
102-06-7	1,3-diphenylguanidine	Suspect
103-26-4	Methyl Cinnamate	Suspect
103-60-6	2-phenoxyethyl isobutyrate	Suspect
104-40-5	4-n-nonylfenol	Top 100
105-75-9	dibutyl fumarate	Suspect
105-76-0	Dibutyl maleate,	Suspect
105-99-7	Adipic acid-di-n-butyl ester	Suspect
10605-21-7	Carbendazim	Top 100
106-14-9	12-hydroxystearic acid,	Suspect
106-20-7	Di-2-ethylhexylamine	Suspect
1066-51-9	Aminomethylphosphonic acid [AMPA]	Top 100
106-79-6	1,8-Octanedicarboxylic acid-bis-methyl ester	Suspect
107-06-2	trans-1,2-dichloroethane	Top 100
1071-83-6	Glyphosate	Top 100
107534-96-3	Tebuconazol	Suspect
108-20-3	diisopropyl ether (DIPE)	Top 100
108-65-6	Propylene glycol 1-methyl ether 2-acetate	Suspect
108-88-3	Methylbenzene [toluene]	Top 100
108-90-7	Chlorobenzene	Top 100
110488-70-5	Dimethomorph (mixture of E + Z)	Suspect
111-15-9	2-Ethoxy ethylacetaat	Suspect
111-20-6	Sebacic acid	Suspect
111-21-7	Ethylenebis-(2-oxyethyl acetate)	Suspect
111-81-9	Methyl-10-undecenoate	Suspect
1120-48-5	Di-n-octylamine	Suspect
112-49-2	Triethyl glycol dimethyl ether [Triglyme]	Suspect

112-50-5	tri(ethylene glycol) monoethyl ether	Suspect
112-75-4	N,N-dimethyl-tetradecylamine	Suspect
114-07-8	erythromycin	Top 100
115-96-8	Tris(2-chloorethyl)fosfaat	Suspect
117-81-7	diethylhexyl phthalate [DEHP]	Top 100
117-82-8	Bis(methylglycol) phthalate	Suspect
117-96-4	Amidotrizoic acid	Top 100
118-60-5	Salicylic acid-2-ethyl-1-hexyl ester	Suspect
1194-65-6	dichlobenil	Top 100
120068-37-3	Fipronil	Suspect
120-36-5	dichloorprop	Top 100
120-71-8	2-Methoxy-5-methylaniline	Suspect
120-83-2	2,4-dichlorophenol	Top 100
122-34-9	simazin	Suspect/Top100
86-66-8	1,3,(6of7)naftaleentrisulfonzuur, triNazout	Top 100
123-99-9	Azelaic acid	Suspect
12645-31-7	Phosphoric acid, 2-ethylhexyl ester	Suspect
126-73-8	Tributyl phosphate	Suspect/Top100
127-18-4	tetrachloroethene	Top 100
128-80-3	Solvent Green 3	Suspect
131-17-9	Ftaalzuur, bis-allylester	Suspect
131341-86-1	Fludioxonil Pestanal	Suspect
13674-84-5	Tris-(2-chloroisopropyl) phosphate	Suspect
137862-53-4	Valsartan	Suspect
138402-11-6	Irbesartan	Suspect
139-13-9	Nitrilotriacetic acid [NTA]	Top 100
140-66-9	4-tert-octylphenol	Top 100
1420-07-1	2-(1,1-dimethylethyl)-4,6-dinitrophenol [dinoterb]	Top 100
143-24-8	Tetraethyleen glycol dimethyl ether	Suspect
143390-89-0	Kresoxim-Methyl	Suspect
60-00-4	Ethyleendiaminetetra-acetic acid [EDTA]	Top 100
150-68-5	monuron	Top 100
150-84-5	Citronellyl acetate	Suspect
15206-55-0	Methyl benzoylformate	Suspect
15307-86-5	Diclofenac	Top 100
15545-48-9	chloortoluron	Top 100
1563-66-2	Carbofuran	Suspect
156-43-4	p-phenetidine	Suspect
156-59-2	cis-1,2-dichloroethene	Top 100
1593-77-7	Dodemorph	Suspect
161326-34-7	Fenamidone	Suspect
1634-04-4	Methyl tertiair-butylether [MTBE]	Top 100
1636-27-7	dipropylmalonic acid	Suspect
1678-25-7	N-phenylbenzenesulphonamide	Suspect

1698-60-8	Chloridazon	Suspect/Top100
17392-83-5	Methyl (R)-(+)-lactate	Suspect
1763-23-1	Heptadecafluorooctanesulfonic acid	Suspect
1852-04-6	undecanedioic acid	Suspect
1912-24-9	Atrazine	Top 100
19937-59-8	metoxuron	Top 100
2008-58-4	BAM	Top 100
21087-64-9	metribuzin	Top 100
23103-98-2	Pirimicarb	Top 100
2327-02-8	1-(3,4-dichloorfenyl)urea	Top 100
2386-87-0	3,4-Epoxycyclohexylmethyl 3,4-epoxycyclohexanecarboxylate	Suspect
2425-79-8	1,4-Butanediol diglycidyl ether	Suspect
2432-99-7	11-aminoundecanoic acid	Suspect
2437-25-4	undecyl cyanide	Suspect
24544-04-5	2,6-diisopropylaniline	Suspect
24748-23-0	3,6,9-Triethyl-3,6,9-trimethyl-1,4,7-triperoxonane	Suspect
25057-89-0	bentazon	Suspect/Top100
25812-30-0	Gemfibrozil	Suspect
2593-15-9	Etridiazole	Top 100
26225-79-6	Ethofumesate	Top 100
2634-33-5	1,2-Benzisothiazol-3(2H)-one	Suspect
2687-94-7	1-octyl-2-pyrrolidinone	Suspect
95-14-7	1H-benzotriazool	Suspect
27871-49-4	(-)-Methyl L-lactate	Suspect
28159-98-0	Irgarol	Suspect
2873-97-4	Diacetone acrylamide	Suspect
288-13-1	pyrazole	
2921-88-2	chloorpyrifos	Top 100
298-00-0	Parathion-Methyl	Top 100
298-46-4	Carbamazepine	Top 100
29878-31-7	4-Methyl-1H-Benzotriazole	Suspect
3006-86-8	1,1-Di(tert-butylperoxy)cyclohexane	Suspect
3060-89-7	metobromuron	Top 100
314-40-9	bromacil	Top 100
3149-12-0	2,6-diethoxytetrahydropyran	Suspect
3195-24-2	diethyl diallylmalonate	Suspect
32210-23-4	4-tert-butylcyclohexyl acetate	Suspect
3290-92-4	trimethylolpropane trimethacrylate	Suspect
330-54-1	Diuron	Top 100
330-55-2	linuron	Top 100
335-67-1	Perfluoroctaanzuur	Suspect
34123-59-6	Isoproturon	Top 100
3567-62-2	1-(3,4-Dichlorophenyl)-3-methylurea	Top 100

3622-84-2	N-n-Butylbenzenesulfonamide	Suspect
36507-30-9	Carbamazepine 10,11-Epoxide	Suspect
37350-58-6	metoprolol	Suspect/Top100
3930-20-9	Sotalol	Top 100
4098-71-9	isophorone diisocyanate	Suspect
41859-67-0	bezafibraat	Top 100
42036-65-7	2-(Dimethylaminomethyl)-1-cyclohexanone hydrochloride	Suspect
4273-98-7	2-Phenylsulfonaniline	Suspect
50940-49-3	MAES	Suspect
51000-52-3	Neodecanoic acid-ethenyl ester	Suspect
51-03-6	Piperonylbutoxide	Suspect
51218-45-2	metolachloor	Suspect/Top100
52722-86-8	4-hydroxy-2,2,6,6-tetramethylpiperidine-1-ethanol	Suspect
534-52-1	DNOC	Top 100
541-02-6	Decamethylcyclopentasiloxane	Suspect
54982-83-1	Musk MC4	Suspect
551-16-6	6-Aminopenicillanic acid	Suspect
551-93-9	2-aminoacetofenon	Top 100
554-00-7	2,4-dichlooraniline	Top 100
5571-36-8	cyclic 3-(1,2-ethanediylacetale)-estra-5(10),9(11)-diene-3,17-dione	Suspect
56-23-5	tetrachloromethane	Top 100
5669-19-2	2-Benzylacrylicacid	Suspect
58-08-2	caffeïne	Suspect/Top100
5888-33-5	isobornyl acrylate	Suspect
5915-41-3	terbutylazin	Suspect/Top100
60166-93-0	iopamidol	Top 100
604-75-1	oxazepam	Suspect
60-80-0	fenazon	Suspect/Top100
61597-98-6	Propanoic acid, 2-hydroxy-, (1R,2S,5R)-5-methyl-2-(1-methylethyl)cyclohexyl ester, (2S)-	Suspect
6190-65-4	Desethylatrazine	Top 100
623-53-0	ethyl methyl carbonate	Suspect
62-53-3	Aniline	Top 100
631-64-1	dibromoacetic acid	Top 100
637-92-3	Ethyl-tertiair-butylether [ETBE]	Top 100
63968-64-9	ARTEMISININ,	Suspect
64744-50-9	4,4-pentamethylene-2-pyrrolidinone	Suspect
6493-05-6	Pentoxifylline	Top 100
6600-31-3	3,9-di-(3-cyclohexenyl)-2,4,8,10-tetraoxaspiro(5,5)undecane	Suspect
66108-95-0	Iohexol	Top 100
67129-08-2	Metazachlor	Top 100
67-43-6	Diethylenetriaminepentaacetic acid [DTPA]	Top 100

67-66-3	Trichloromethane [chloroform]	Top 100
688-84-6	methacrylic acid-2-ethylhexyl ester	Suspect
6938-94-9	diisopropyl adipate	Suspect
704-00-7	1,2-diacetylbenzene	Suspect
7085-19-0	mecoprop (MCPP)	Suspect/Top100
111-96-6	diethylene glycol dimethyl ether [diglyme]	
71-43-2	Benzene	Top 100
71-55-6	1,1,1-trichloroethane	
723-46-6	sulfametoxazool	Top 100
7328-22-5	diethylene glycol butyl ether	Suspect
73334-07-3	Iopromide	Top 100
73942-87-7	7,8-dimethoxy-1,3-dihydro-2H-3-benzazepin-2-one	Suspect
7397-62-8	butyl glycolate	Suspect
7491-09-0	Docusate Potassium	Suspect
75-09-2	dichloromethane	Top 100
75-25-2	tribromomethane = bromoform	Top 100
75-27-4	bromodichloromethane	Top 100
7534-94-3	isobornyl methacrylate	Suspect
7547-66-2	2,4-dichlorophenoxyacetic acid [2,4-D]	Top 100
76-03-9	trichloroacetic acid [TCA]	Top 100
77-93-0	triethyl citrate	Suspect
78-08-0	Triethoxyvinylsilane	Suspect
78-40-0	triethyl phosphate	Suspect
78-63-7	2,5-Dimethyl-2,5-di(tert-butylperoxy)hexane	Suspect
79-01-6	trichloroethene	Top 100
79-11-8	monochloroacetic acid	Top 100
791-28-6	TPPO	Suspect/Top100
79-43-6	dichloroacetic acid	Top 100
80-05-7	Bisphenol-A	Top 100
80-09-1	4,4'-sulphonyldiphenol = bisfenol-S	Suspect/Top100
81-14-1	Musk Ketone	Suspect
81405-85-8	etridiazool	Suspect
826-36-8	2,2,6,6-tetramethyl-4-piperidone	Suspect
826-81-3	8-hydroxyquinaldine	Suspect
83-15-8	4-Acetamidoantipyrine	Suspect
84-66-2	diethyl phthalate [DEPH]	Top 100
84-69-5	phthalic acid, bis-iso-butyl ester	Suspect
84-74-2	phthalic acid, bis-n-butyl ester	Suspect/Top100
68153-01-5	naftaleen-1-sulfonzuur	Top 100
85-98-3	1,3-Diethyl-1,3 diphenylurea	Suspect
86-50-0	azinfos-methyl	Top 100
88-06-2	2,4,6-trichloorfenol	Top 100
88671-89-0	myclobutanil	Suspect
89-48-5	menthyl Acetaat	Suspect

90315-82-5	Ethyl (R)-2-hydroxy-4-phenylbutyrate,	Suspect
90-98-2	4,4'-dichlorobenzophenone	Suspect
91-20-3	naphthalene	Top 100
924-88-9	diisopropyl succinate	Suspect
94-70-2	o-phenetidine	Suspect
94-74-6	4-chloro-2-methylphenoxyacetic acid [MCPA]	Suspect/Top100
95-50-1	1,2-dichlorobenzene	Top 100
96562-58-2	methyl (r)-2-(4-hydroxyphenoxy)propionate	Suspect
97-78-9	N-lauroylsarcosine	Suspect
98-82-8	isopropylbenzene [Cumol]	Top 100
98967-40-9	flumetsulam	Suspect
13194-48-4	ethoprofos	
94-59-7	safrole	Suspect
50892-62-1	8-chloro-5,10-dihydro-11H-dibenzo[b,e][1,4]diazepin-11-one	Suspect
56718-70-8	1-[4-(2-Methoxyethyl)phenoxy]-2,3-epoxypropane	Suspect
97963-62-7	5-(difluoromethoxy)-1H-benzimidazole-2-thiol	Suspect

Attachment III Vitens list of studied target compounds

CAS number	compound name
6493-05-6	Pentoxifylline
8055-08-1	Acetaminophen
100646-51-3	Quizalofop-P-Ethyl
1007-28-9	Atrazinee-desisopropyl
100-88-9	Cyclamate
102-65-8	Sulfachloropyridazine
102962-29-8	Diuron
10309-95-2	Malachite green
104206-82-8	Mesotrione
104206-82-8	Mesotrione
104732-42-5	3-iodo-2-propynyl N-butylcarbamate
10540-29-1	Tamoxifen
105512-06-9	Clodinafop-propargyl
107534-96-3	Tebuconazole
110235-47-7	Mepanipyrim
110488-70-5	Dimethomorph
11096-88-1	Cyanazine
11111-56-1	Chlortoluron
1113-02-6	Omethoate
111988-49-9	Thiacloprid
111991-09-4	Nicosulfuron
112143-77-8	Chlorsulfuron
114-07-8	Erythromycin
114-26-1	Propoxur
114798-26-4	Losartan
116-06-3	Aldicarb
119168-77-3	Tebufenpyrad
119446-68-3	Difenoconazole
119603-94-0	Simazine
120068-37-3	Fipronil
120116-88-3	Cyazofamid
120162-55-2	Azimsulfuron
120-36-5	Dichlorprop
120923-37-7	Amidosulfuron
121552-61-2	Cyprodinil

122-11-2	Sulfadimethoxin
122-14-5	Fenitrothion
122667-23-6	2-octyl-4-isothiazoline-3-one
122-88-3	4-CPA
122931-48-0	Rimsulfuron
123113-74-6	Acetochlor
123312-89-0	Pymetrozine
125116-23-6	Metconazole
125-33-7	Primidone
12542-35-7	Propyphenazone
126535-15-7	Triflusulfuron-methyl
126833-17-8	Fenhexamid
127-79-7	Sulfamerazine
128639-02-1	Carfentrazone-ethyl
129378-89-8	Sulfamethoxazole
13013-17-7	Propranolol
131860-33-8	Azoxystrobin
13360-45-7	Chlorbromuron
133855-98-8	Epoxiconazole
13523-86-9	Pindolol
13684-56-5	Desmedipham
13684-63-4	Phenmedipham
136-95-8	2-aminobenzothiazole
137-58-6	Lidocaine
137662-59-0	5-chloro-2-methyl-4-isothiazolin-3-one
137862-53-4	Valsartan
138261-41-3	Imidacloprid
139-40-2	Propazine
1401-69-0	Tylosin
141776-32-1	Sulfosulfuron
1420-07-1	Dinoterb
142459-58-3	Flufenacet
143390-89-0	Kresoxim-methyl
143984-63-8	Flumequine
144651-06-9	Oxasulfuron
144-83-2	Sulfapyridine
14698-29-4	Oxolinic acid
148-79-8	Thiabendazole
1491-59-4	Oxymetazoline
150-68-5	Monuron
153012-39-6	Cefuroxime
15307-86-5	Diclofenac
154-21-2	Lincomycin
154361-50-9	Capecitabine

1563-66-2	Carbofuran
158062-67-0	Flonicamid
15972-60-8	Alachlor
16118-49-3	Carbetamide
1646-88-4	Aldicarb-sulfone
16655-82-6	3-Hydroxycarbofuran
1689-84-5	Bromoxynil
1698-60-8	Chloridazone
172964-50-0	Ketoprofen
173159-57-4	Foramsulfuron
173159-57-4	Foramsulfuron
173584-44-6	Indoxacarb
1746-81-2	Monolinuron
175013-18-0	Pyraclostrobin
1763-23-1	PFOS
18559-94-9	Salbutamol
188425-85-6	Boscalid
1912-24-9	Atrazinee
1912-26-1	Trietazine
1918-00-9	Dicamba
1918-16-7	Propachlor
1929-88-0	Benzthiazuron
1951-25-3	Amiodarone
196618-13-0	Oseltamivir
19928-35-9	Methomyl
19937-59-8	Metoxuron
2032-65-7	Methiocarb
208465-21-8	Mesosulfuron-Methyl
21087-64-9	Metribuzin
21312-10-7	Acetylsulfamethoxazole
2164-17-2	Fluometuron
22224-92-6	Fenamiphos
23135-22-0	Oxamyl
23560-59-0	Heptenophos
23783-98-4	Phosphamidon
23950-58-5	Propyzamide
24017-47-8	Triazophos
243973-20-8	Pinoxaden
25057-89-0	Bentazone
25059-80-7	Benazolin-ethyl
25812-30-0	Gemfibrozil
26159-31-9	Naproxen
26225-79-6	Ethofumesate
26787-78-0	Amoxicillin

27948-47-6	Sotalol
28159-98-0	Irgarol
29232-93-7	Pirimiphos-methyl
298-03-3	Demeton-O
298-46-4	Carbamazepine
29973-13-5	Ethiofencarb
3060-89-7	Metobromuron
311-45-5	Paraoxon-ethyl
31431-39-7	Mebendazole
314-40-9	Bromacil
31879-05-7	Fenoprofen
3337-71-1	Asulam
335-67-1	PFOA
34123-59-6	Isoproturon
34681-10-2	Butocarboxim
34681-23-7	Butoxycarboxim
35554-44-0	Imazalil
361377-29-9	Fluoxastrobin
36341-88-5	Ifosfamid
37350-58-6	Metoprolol
38260-54-7	Etrimfos
39403-80-0	Dinoseb
39410-70-3	Tolyltriazole
40487-42-1	Pendimethalin
41394-05-2	Metamitron
41483-43-6	Bupirimate
41859-67-0	Bezafibrate
4433-52-7	Benzothiazolin
46719-29-3	Terbutaline
49562-28-9	Fenofibrate
50-18-0	Cyclophosphamide
50499-60-0	Clenbuterol
51235-04-2	Hexazinone
51274-03-4	Carbaryl
51-28-5	2,4-Dinitrophenol
5250-39-5	Flucloxacillin
52-68-6	Trichlorfon
52888-80-9	Prosulfocarb
53112-28-0	Pyrimethanil
53-16-7	Estrone
53240-95-2	DNOC
53-86-1	Indometacin
53906-69-7	Aspartame
53906-69-7	Aspartame

54-31-9	Furosemide
55179-31-2	Bitertanol
551-92-8	Dimetridazole
55219-65-3	Triadimenol
55297-95-5	Tiamulin
55335-06-3	Triclopyr
55-38-9	Fenthion
55512-33-9	Pyridate
55762-76-0	Metolachlor
56038-13-2	Sucralose
56645-87-5	Linuron
57226-07-0	Fluoxetine
57646-30-7	Furalaxyl
57-68-1	Sulfadimidine
57837-19-1	Metalaxyl
579-51-1	Chloramphenicol
57966-95-7	Cymoxanil
58560-75-1	Ibuprofen
5915-41-3	Terbutylazine
59-40-5	Sulfaquinoxaline
60207-90-1	Propiconazole
60-54-8	Tetracycline
60-80-0	Phenazone
60966-51-0	Atenolol
615-22-5	2-(methylthio)benzothiazole
6153-64-6	Oxytetracycline
61-72-3	Cloxacillin
6190-65-4	Atrazinee-desethyl
62-44-2	Phenacetin
62883-00-5	Iopamidol
63278-70-6	Carbendazim
63659-18-7	Betaxolol
637-07-0	Clofibrate
66063-05-6	Pencycuron
66108-95-0	Iohexol
66246-88-6	Penconazole
66332-96-5	Flutolanil
66-79-5	Oxacillin
66-79-5	Oxacillin
67306-00-7	Fenpropidin
67306-03-0	Fenpropimorph
67485-29-4	Hydramethylnon
67747-09-5	Prochloraz
68-35-9	Sulfadiazine

68694-11-1	Triflumizol
69377-81-7	Fluroxypyr
69806-50-4	Fluazifop-butyl
70458-96-7	Norfloxacin
7085-19-0	Mecoprop
71701-02-5	Caffeine
7220-97-5	Chlortetracycline
72558-82-8	Ceftazidime
7286-69-3	Sebuthylazine
7287-19-6	Prometryn
73334-07-3	Iopromide
738-70-5	Trimethoprim
74011-58-8	Enoxacin
74223-64-6	Metsulfuron-methyl
75847-73-3	Enalapril
7681-76-7	Ronidazole
77732-09-3	Oxadixyl
78649-41-9	Iomeprol
791-28-6	Triphenylphosphine oxide
79277-27-3	Thifensulfuron-methyl
79622-59-6	Fluazinam
79902-63-9	Simvastatin
80-08-0	Dapson
81-07-2	Saccharin
81103-11-9	Clarithromycin
81777-89-1	Clomazone
82097-50-5	Triasulfuron
83164-33-4	Diflufenican
83380-47-6	Ofloxacin
83905-01-5	Azithromycin
85721-33-1	Ciprofloxacin
87130-20-9	Diethofencarb
87674-68-8	Dimethenamid
87714-45-2	Fenoxaprop-ethyl
882-09-7	Clofibric acid
88-75-5	2-Nitrophenol
90717-03-6	Quinmerac
919-86-8	Demeton-S-methyl
93106-60-6	Enrofloxacin
93-72-1	Fenoprop
94125-34-5	Prosulfuron
94271-03-1	DEET
94361-06-5	Cyproconazole
94-74-6	MCPA

94-81-5	MCPB
94-82-6	2,4-DB
950-35-6	Paraoxon-methyl
96-83-3	Iopanoic acid
98886-44-3	Fosthiazate
99105-77-8	Sulcotrione
99607-70-2	Cloquintocet-mexyl
99616-64-5	Metronidazole

Attachment IV List of compounds relevant to drinking water

CAS number	Compound name	Use
3013-02-3	MTDC	pharmaceutical
100-42-5	Benzene, ethenyl- [styrene]	industrial
100-97-0	Hexamine [urotropin]	industrial
102-06-7	Guanidine, N,N -diphenyl-	industrial
103-90-2	Acetamide, N-(4-hydroxyphenyl)-	pharmaceutical
10605-21-7	Carbendazim	pesticide
106-42-3	Benzene, 1,4-dimethyl-	industrial
106-47-8	Benzenamine, 4-chloro-	industrial
1066-51-9	Aminomethylphosphonic acid [AMPA]	pesticide
107-06-2	Ethane, 1,2-dichloro-	industrial
1071-83-6	Glycine, N-(phosphonomethyl)- [glyphosate]	pesticide
107534-96-3	Tebuconazole	pesticide
108-20-3	Propane, 2,2 -oxybis-	industrial
108-38-3	Benzene, 1,3-dimethyl-	industrial
108-88-3	Benzene, methyl- [toluene]	industrial
108-95-2	Phenol	industrial
109-87-5	Methane, dimethoxy-	industrial
109-99-9	Furan, tetrahydro-	industrial
110488-70-5	Dimethomorph	pesticide
110-82-7	Cyclohexane	industrial
111-96-6	Ethane, 1,1 -oxybis[2-methoxy- [diglyme]	industrial
111991-09-4	Nicosulfuron	pesticide
112-49-2	2,5,8,11-Tetraoxadodecane [triglyme]	industrial
117-96-4	Benzoic acid, 3,5-bis(acetylamino)-2,4,6-triiodo- [amidotrizoic acid]	contrast medium
120-12-7	Anthracene	industrial
120-46-7	1,3-Propanedione, 1,3-diphenyl-	industrial
120-78-5	Benzothiazole, 2,2 -dithiobis-	industrial
120-82-1	Benzene, 1,2,4-trichloro-	industrial
122-34-9	Simazine	pesticide
122-88-3	Acetic acid, (4-chlorophenoxy)-	pesticide
123-91-1	1,4-Dioxane	industrial
124-48-1	Methane, dibromochloro-	industrial
125-33-7	Primidone	pharmaceutical
126-71-6	Phosphoric acid, tris(2-methylpropyl) ester	industrial
126-73-8	Phosphoric acid tributyl ester	industrial

127-18-4	Ethene, tetrachloro-	industrial
133-07-3	Folpet	pesticide
13429-07-7	2-Propanol, 1-(2-methoxypropoxy)-	industrial
134-62-3	Diethyl toluamide [DEET]	pesticide
13674-84-5	2-Propanol, 1-chloro-, phosphate (3:1)	industrial
136-85-6	1H-Benzotriazole, 5-methyl-	industrial
139-13-9	Glycine, N,N-bis(carboxymethyl)- [nitrolotriacetic acid]	industire
1420-07-1	Dinoterb	pesticide
143-24-8	2,5,8,11,14-Pentaoxapentadecane [tetraglyme]	industrial
148-79-8	Thiabendazole	pesticide
149-30-4	2(3H)-Benzothiazolethione	pesticide
15045-43-9	2,2,5,5-Tetramethyl-tetrahydrofuran	industrial
152019-73-3	Metolachlor OA	metabolite
15307-86-5	Diclofenac	pharmaceutical
156-59-2	Ethene, 1,2-dichloro-, (Z)-	industrial
156-60-5	Ethene, 1,2-dichloro-, (E)-	industrial
15972-60-8	Alachlor	pesticide
1634-04-4	Ether, methyl tert-butyl [MTBE}	industrial
163515-14-8	Dimethenamid-P	pesticide
1646-87-3	Aldicarb sulfoxide	pesticide
1698-60-8	Chloridazon	pesticide
171118-09-5	Metolachlor ESA	metabolite
17254-80-7	Desphenylchloridazon, methyl-	metabolite
187022-11-3	Acetochlor ESA sodium salt	pesticide
1918-16-7	Propachlor	pesticide
2008-58-4	2,6-Dichlorobenzamide	pesticide
205939-58-8	Dimethenamid ESA	metabolite
22071-15-4	Ketoprofen	pharmaceutical
23135-22-0	Oxamyl	pesticide
2371-42-8	exo-1,2,7,7-Tetramethylbicyclo[2.2.1]heptan-2-ol	natural origin
2387-23-7	Urea, N,N'-dicyclohexyl-	pharmaceutical
24579-73-5	Propamocarb	pesticide
25057-89-0	Bentazone	pesticide
25812-30-0	Gemfibrozil	pharmaceutical
2593-15-9	Ethazole	pesticide
2634-33-5	1,2-Benzisothiazol-3(2H)-one	pesticide
27203-92-5	Tramadol	pharmaceutical
28179-44-4	Ioxitalamic acid	contrast medium
29385-43-1	1H-Benzotriazole, 4(or 5)-methyl-	industrial
298-46-4	Carbamazepine	pharmaceutical
29878-31-7	1H-Benzotriazole, 4-methyl-	industrial
30391-89-0	Benzamide, 2-amino-N-(1-methylethyl)-	metabolite
304-55-2	Butanedioic acid, 2,3-dimercapto-, (R*,S*)-	pharmaceutical
314-40-9	2,4(1H,3H)-Pyrimidinedione, 5-bromo-6-methyl-3-(1-methylpropyl)-	pesticide

	[bromacil]	
31879-05-7	Fenoprofen	pharmaceutical
330-54-1	Diuron	pesticide
330-55-2	Linuron	pesticide
338-45-4	Mevinphos, trans-isomer	pesticide
34123-59-6	Isoproturon	pesticide
34681-10-2	Butocarboxim	pesticide
34681-23-7	Butoxycarboxim	pesticide
36507-30-9	Carbamazepine 10,11-epoxide	metabolite
37350-58-6	Metoprolol	pharmaceutical
39184-27-5	Thiofanox sulfoxide	pesticide
3984-14-3	Sulfamide, N,N-dimethyl-	metabolite
4184-79-6	1H-Benzotriazole, 5,6-dimethyl-	industrial
45951-45-9	Sulfamic acid, N-cyclohexyl-	food ingredient
479-92-5	Propyphenazone	pharmaceutical
49562-28-9	Fenofibrate	pharmaceutical
496-11-7	1H-Indene, 2,3-dihydro-	industrial
50-78-2	Aspirin	pharmaceutical
51218-45-2	Metolachlor	pesticide
52508-35-7	Dikegulac sodium	pesticide
525-66-6	Propranol	pharmaceutical
53-16-7	Estrone	natural origin
54-31-9	Furosemide	pharmaceutical
55297-95-5	Tiamulin	pharmaceutical
55589-62-3	Acesulfame-K	food ingredient
56038-13-2	Sucralose	food ingredient
56-65-5	Adenosine 5 -(tetrahydrogen triphosphate) [ATP]	natural origin
57-62-5	Chlortetracycline	pharmaceutical
57-68-1	Sulfadimidine	pharmaceutical
58-08-2	Caffeine	food ingredient
58-55-9	Theophylline	pharmaceutical
58-93-5	Hydrochlorothiazide	pharmaceutical
59017-64-0	Ioxaglic acid	contrast medium
5915-41-3	Terbutylazine [TBA}	pesticide
60-00-4	Ethylenediaminetetraacetic acid [EDTA]	industrial
604-75-1	Oxazepam	pharmaceutical
60-80-0	Phenazone	pharmaceutical
61-33-6	Pencillin G	pharmaceutical
61-56-3	Sulthiame	pharmaceutical
61869-08-7	Paroxetine	pharmaceutical
62-53-3	Benzenamine [aniline]	industrial
62-75-9	Methanamine, N-methyl-N-nitroso- [NDMA]	industrial
62883-00-5	Iopamidol	contrast medium
631-64-1	Acetic acid, dibromo-	pharmaceutical

63-25-2	Carbaryl	pesticide
6339-19-1	Desphenylchloridazon	metabolite
657-24-9	Metformin	pharmaceutical
66108-95-0	Iohexol	contrast medium
67-43-6	Glycine, N,N-bis 2- bis(carboxymethyl)amino ethyl -	industrial
67-64-1	Acetone	industrial
67-66-3	Methane, trichloro- [chloroform]	industrial
67-72-1	Ethane, hexachloro-	industrial
68002-20-0	Melamine, hexa(methoxymethyl)-	industrial
69-72-7	Benzoic acid, 2-hydroxy-	pharmaceutical
7085-19-0	Mecoprop (racemate)	pesticide
71-55-6	Ethane, 1,1,1-trichloro-	industrial
73334-07-3	Iopromide	contrast medium
74-95-3	Methane, dibromo-	industrial
75-01-4	Ethene, chloro- [vinyl chloride]	industrial
75-09-2	Methane, dichloro-	industrial
75-25-2	Methane, tribromo- [bromoform]	industrial
75-27-4	Methane, bromodichloro-	industrial
75-35-4	Ethene, 1,1-dichloro-	industrial
75-62-7	Methane, bromotrichloro-	industrial
76-03-9	Acetic acid, trichloro-	industrial
78-40-0	Phosphoric acid, triethyl ester	industrial
78649-41-9	Iomeprol	contrast medium
78-87-5	Propane, 1,2-dichloro-	industrial
79-00-5	Ethane, 1,1,2-trichloro-	industrial
79-01-6	Ethene, trichloro-	industrial
79-11-8	Acetic acid, chloro-	industrial
791-28-6	Phosphine oxide, triphenyl- [TPPO]	industrial
79-43-6	Acetic acid, dichloro-	industrial
83-15-8	4-Acetamidoantipyrin	metabolite
84-66-2	Diethyl phthalate	industrial
882-09-7	Clofibric acid	pesticide
91-20-3	Naphthalene	industrial
93-65-2	Propanoic acid, 2-(4-chloro-2-methylphenoxy)-	pesticide
94-74-6	Acetic acid, (4-chloro-2-methylphenoxy)- [MCPA]	pesticide
95-14-7	1H-Benzotriazole	industrial
95-50-1	Benzene, 1,2-dichloro-	industrial
95-51-2	Benzenamine, 2-chloro-	industrial
96-18-4	Propane, 1,2,3-trichloro-	industrial
99105-77-8	Sulcotrione	pesticide
Multiple	Gadolinium compounds	contrast medium

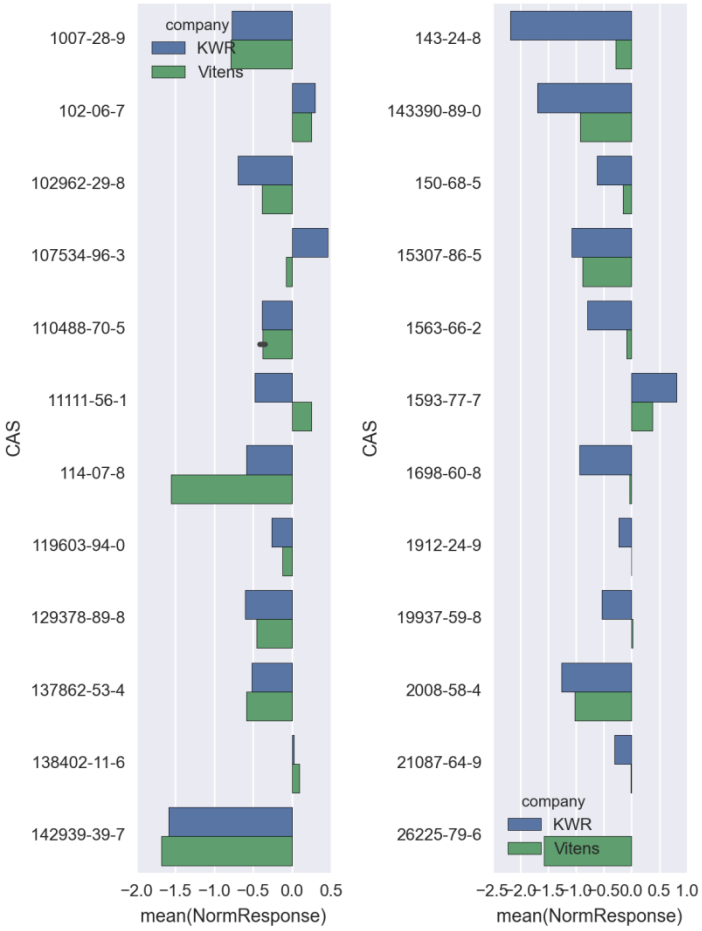
Attachment V: compared response of Vitens and KWR

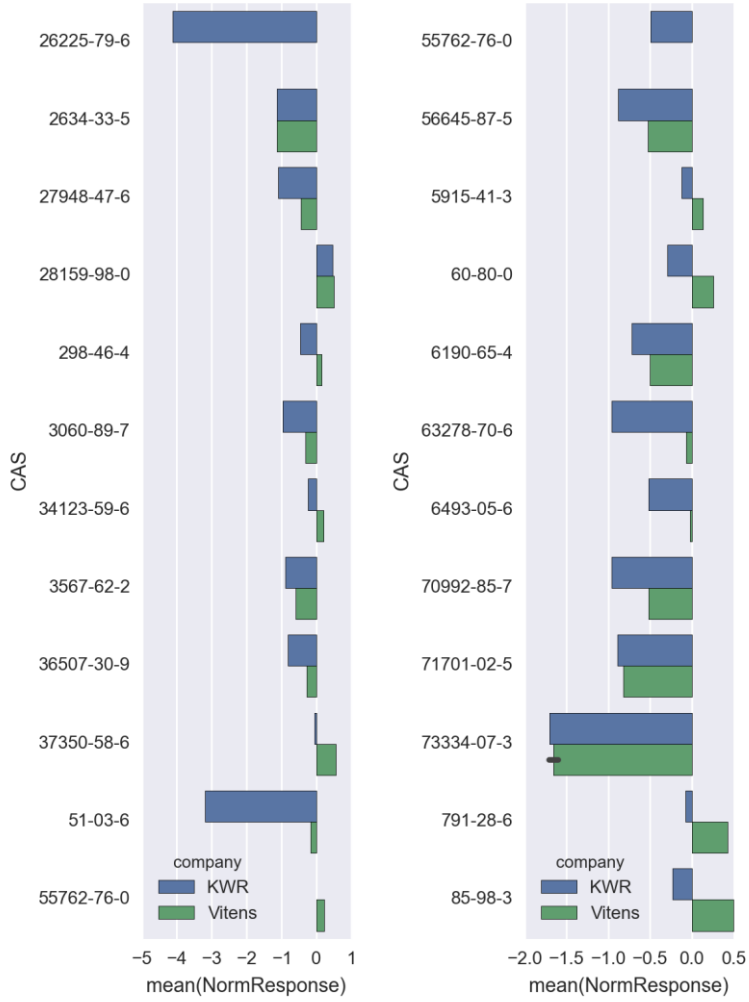
Response factors in the positive ionisation mode (atrazine-d5 eq.) in the LC-HRMS screening by Vitens and KWR.

Response factors in the positive ionisation mode (atrazine-d5 eq.)	Peak area relative to internal standard		Peak area relative to the average peak area	
Component	Vitens	KWR	Vitens	KWR
1-(3.4-dichlorophenyl)-3-methylurea	0.252	0.128	0.299	0.141
1.2-Benzothiazolin-3-one	0.074	0.076	0.088	0.136
1.3-Diethyl-1.3-diphenylurea	3.144	0.584	3.738	1.217
1.3-diphenylguanidine	1.768	1.974	2.103	3.610
2.6-Dichlorobenzamide (BAM)	0.095	0.055	0.113	0.065
Atrazinee	1.000	0.586	1.189	0.854
Atrazinee-desethyl	0.313	0.191	0.372	0.310
Atrazinee-desisopropyl	0.163	0.168	0.194	0.207
Bezafibrate	0.357	0.000	0.425	0.000
Caffeine	0.150	0.130	0.179	0.223
Carbamazepine	1.444	0.344	1.718	0.569
Carbamazepine 10.11-epoxide	0.529	0.156	0.629	0.420
Carbendazim	0.857	0.109	1.019	0.139
Carbofuran	0.802	0.162	0.954	0.330
Chloridazone	0.914	0.116	1.086	0.127
Chlortoluron	1.763	0.329	2.096	0.395
Diclofenac	0.130	0.083	0.155	0.145
Di-glyme	0.021	0.026	0.050	0.039
Dimethomorph-A	0.382	0.406	0.455	0.977
Dimethomorph-B	0.452	0.406	0.537	0.977
Diuron	0.407	0.201	0.484	0.277
Dodemorph-A	2.373	6.476	2.822	9.567
Dodemorph-B	2.379	6.476	2.828	9.567
Erythromycin	0.028	0.259	0.066	0.281
Ethofumesate	0.027	0.000	0.064	0.000
Iohexol	0.026	0.000	0.062	0.000
Iopamidol	0.024	0.000	0.057	0.000
Iopromide	0.025	0.020	0.059	0.030
Iopromide-01	0.020	0.020	0.047	0.030
Irbesartan	1.235	1.062	1.468	1.899
Irgarol	3.136	2.960	3.729	5.302
Isoproturon	1.603	0.588	1.906	0.766
Kresoxim-methyl	0.120	0.020	0.284	0.044
Linuron	0.298	0.131	0.354	0.148
Metobromuron	0.485	0.107	0.577	0.126
Metolachlor	1.659	0.317	1.973	0.580

Response factors in the positive ionisation mode (atrazine-d5 eq.)	Peak area relative to internal standard		Peak area relative to the average peak area	
Component	**Vitens**	**KWR**	**Vitens**	**KWR**
Metoprolol	3.703	0.874	0.883	1.484
Metoxuron	1.039	0.290	1.236	0.361
Metribuzin	0.975	0.492	1.158	0.619
Monuron	0.704	0.240	0.838	0.354
Pentoxifylline	0.954	0.301	1.134	0.318
Phenazone	1.802	0.508	2.142	0.604
Piperonyl-butoxide	0.685	0.001	1.629	0.002
Simazine	0.755	0.545	0.898	0.649
Sotalol	0.363	0.079	0.432	0.089
Sulfamethoxazole	0.350	0.248	0.415	0.303
Tebuconazole	0.839	2.852	0.998	4.665
Terbutylazine	1.347	0.750	1.602	0.979
Tetra-glyme	0.518	0.007	0.616	0.013
Tri-glyme	0.302	0.109	0.360	0.133
Triphenylphosphine oxide	2.671	0.835	3.175	1.400
Valsartan	0.256	0.300	0.304	0.531

Normalised (relative to internal standard atrazine-d5) and log transformed response compared for Vitens and KWR; LC-HRMS screening in the positive ionisation mode.

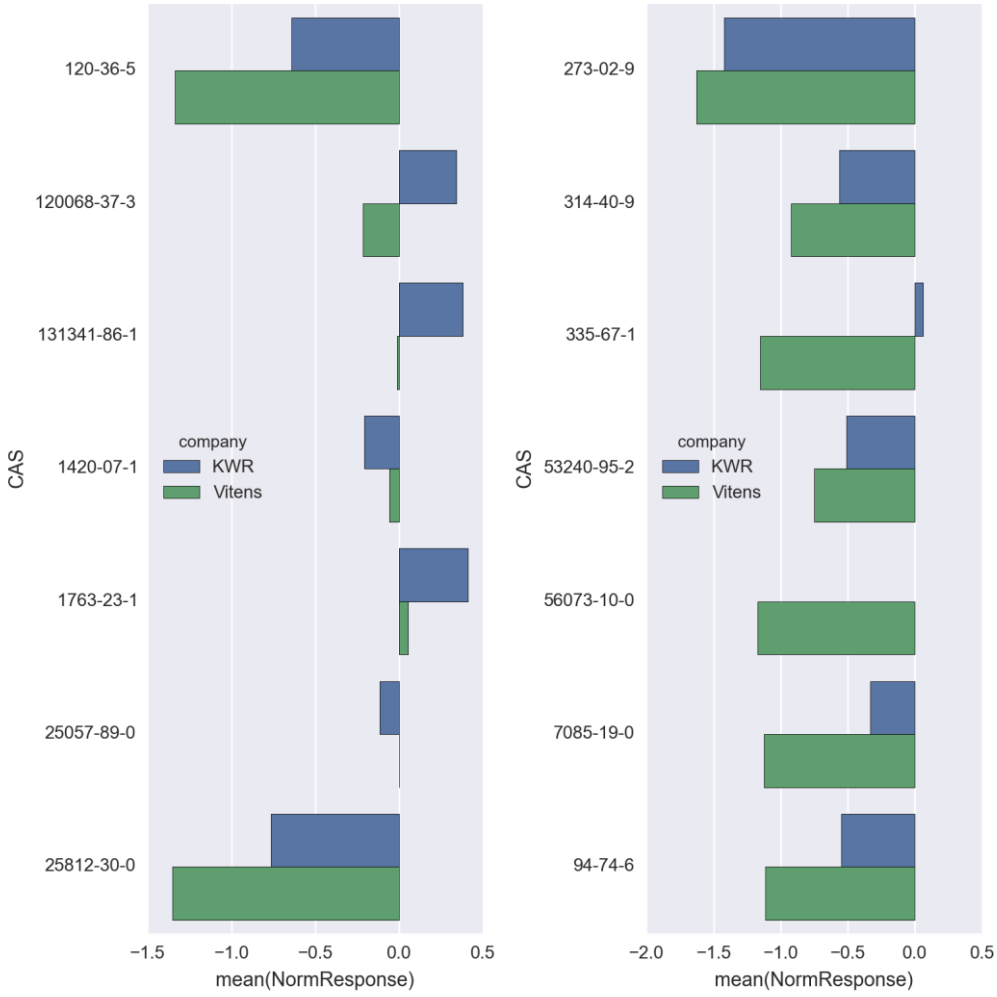



26225-79-6
2634-33-5
27948-47-6
28159-98-0
298-46-4
3060-89-7
34123-59-6
3567-62-2
36507-30-9
37350-58-6
51-03-6
55762-76-0
CAS
company
KWR
Vitens
−5 −4 −3 −2 −1 0 1
mean(NormResponse)
55762-76-0
56645-87-5
5915-41-3
60-80-0
6190-65-4
63278-70-6
6493-05-6
70992-85-7
71701-02-5
73334-07-3
791-28-6
85-98-3
CAS
company
KWR
Vitens
−2.0 −1.5 −1.0 −0.5 0.0 0.5
mean(NormResponse)

Response factors in the negative ionisation mode (bentazon-d6 eq.) for the LC-HRMS screening by Vitens and KWR.

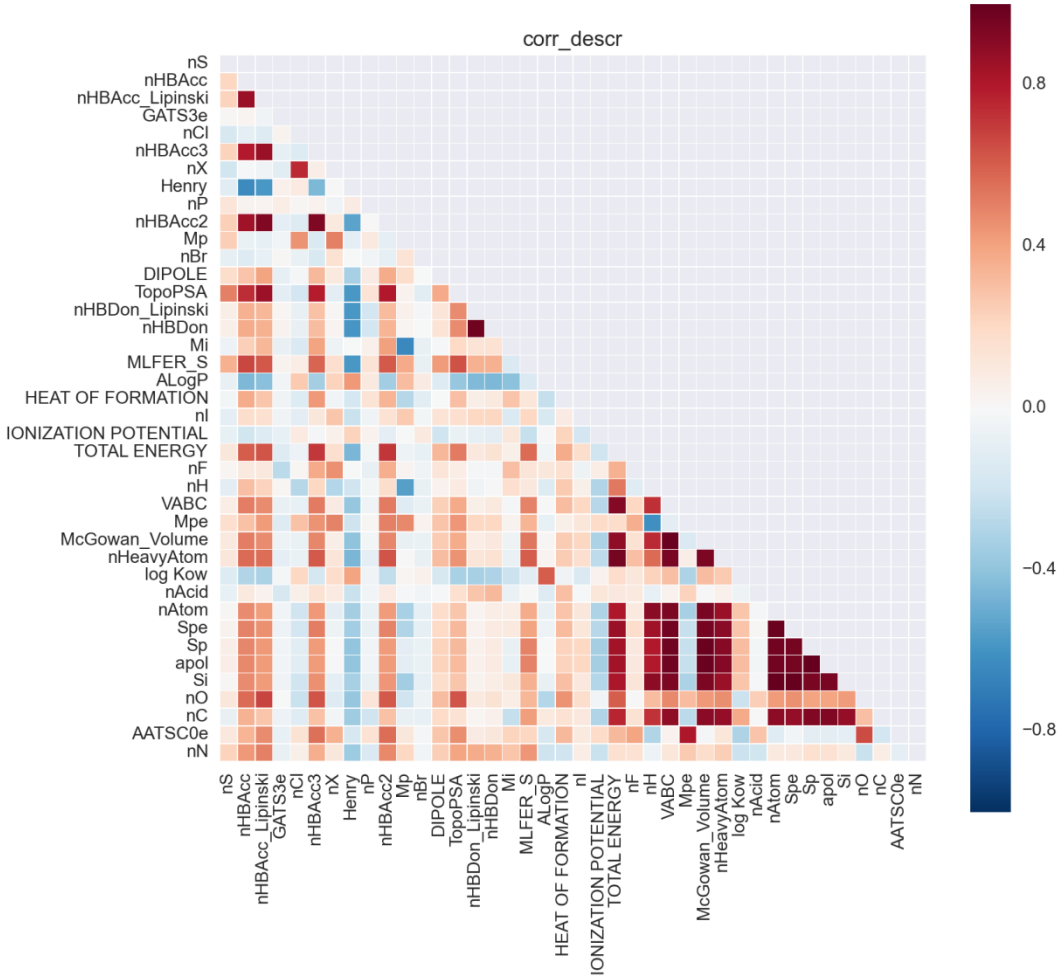
Response factors in the negative ionisation mode (bentazon-d6 eq.)	Peak area relative to the internal standard		Peak area relative to the average peak area	
Component	**Vitens**	**KWR**	**Vitens**	**KWR**
Bentazone	1.000	0.766	2.196	0.535
Bromacil	0.119	0.274	0.261	0.243
Dichlorprop	0.046	0.000	0.101	0.000
Dinoterb	0.874	0.621	1.918	0.532
Fipronil	0.609	2.189	1.337	2.647
Fludioxonil	0.970	2.390	2.131	2.931
Gemfibrozil	0.044	0.173	0.098	0.223
MCPA	0.076	0.281	0.168	0.303
Mecoprop (MCPP)	0.075	0.466	0.164	0.432
PFOA	0.070	1.146	0.154	1.204
PFOS	1.126	2.562	2.472	2.486

Figure: Normalised (relative to internal standard bentazon-d6) and log transformed response compared for Vitens and KWR; LC-HRMS screening in the negative ionisation mode

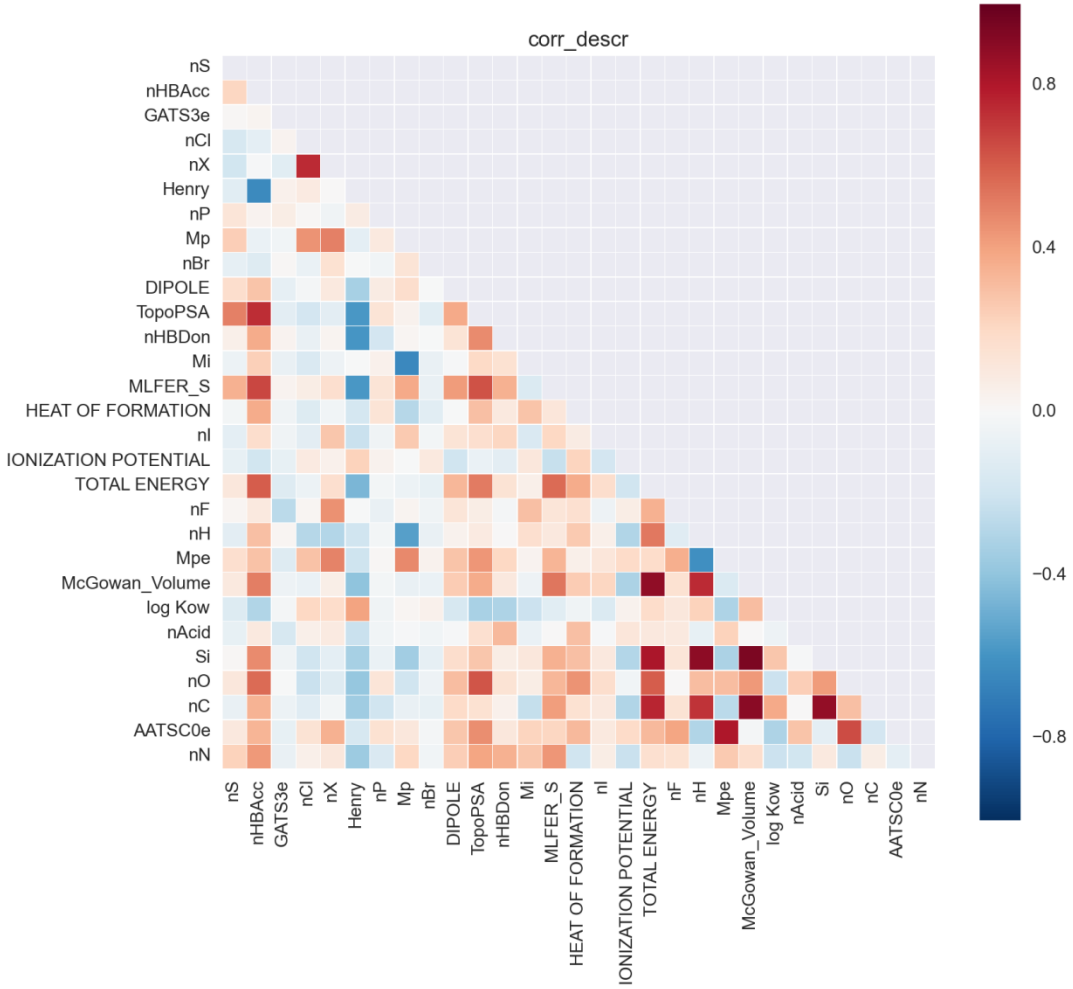


Attachment VI Correlation of the different descriptor values

Starting set of descriptors:



Descriptor set after removing strongly related descriptors:

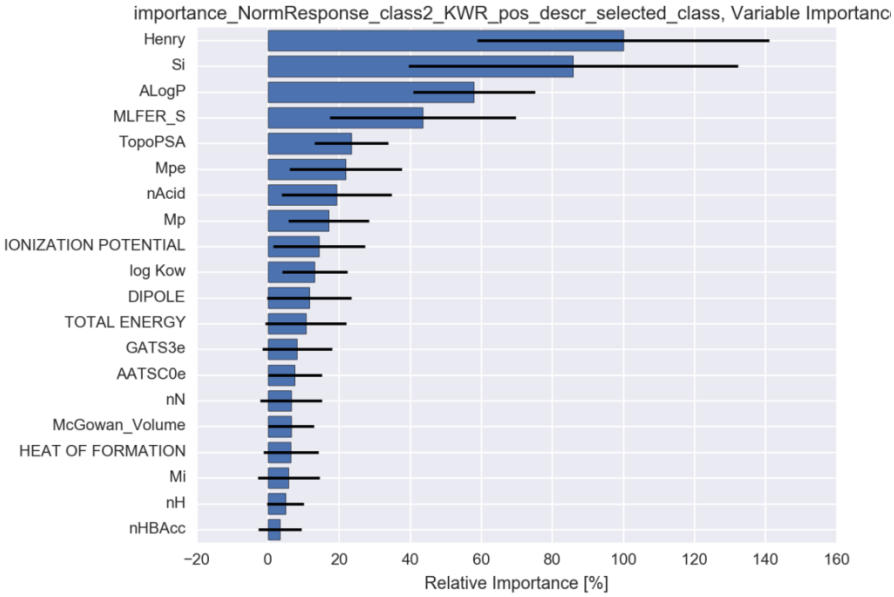


R^2 of descriptors to response in six datasets:

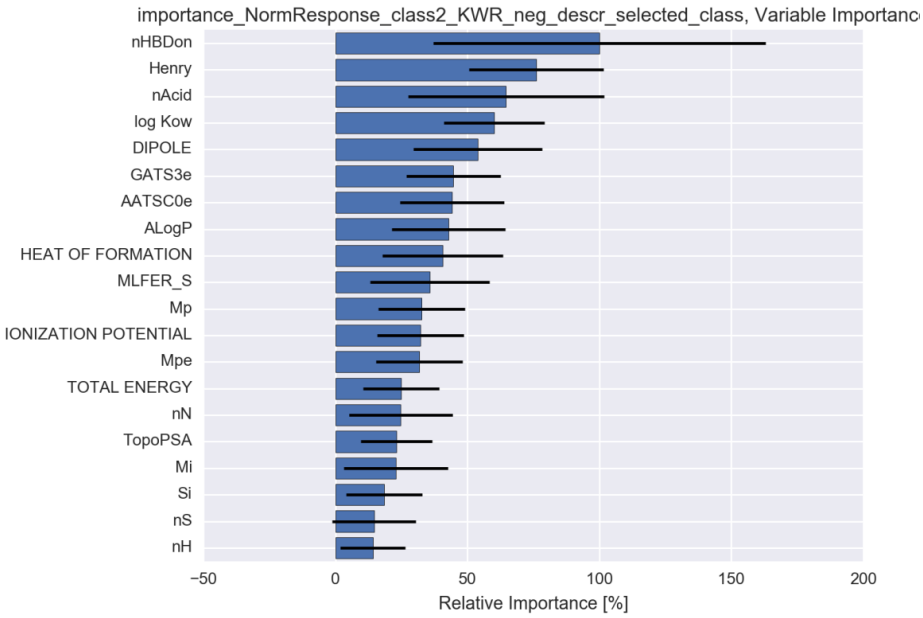
Company	KWR	KWR	Vitens	KWR	KWR	Vitens
Ionisation mode	neg	neg	neg	pos	pos	pos
Number of compounds	67	65	44	135	132	259
LC column	Without LC	With LC	With LC	Without LC	With LC	With LC
AATSC0e	0.22	0.22	0.24	-0.60	-0.47	-0.41
DIPOLE	0.12	0.13	0.28	0.02	0.07	-0.20
GATS3e	-0.47	-0.47	-0.40	0.13	0.08	0.16
HEAT OF FORMATION	0.17	0.24	-0.16	-0.44	-0.38	-0.06
Henry	0.05	0.09	0.14	0.04	-0.13	0.19
IONISATION POTENTIAL	0.35	0.38	0.42	-0.33	-0.32	-0.07
MLFER_S	-0.12	-0.09	-0.12	0.06	0.21	-0.13
McGowan_Volume	-0.03	0.04	0.02	-0.02	0.12	0.10
Mi	0.19	0.18	0.42	0.01	-0.01	0.08
Mp	-0.16	-0.10	-0.23	0.15	0.21	-0.25
Mpe	0.15	0.13	0.29	-0.26	-0.17	-0.42
Si	0.02	0.06	0.04	-0.03	0.08	0.21
TOTAL ENERGY	0.20	0.26	0.23	-0.21	-0.03	-0.01
TopoPSA	0.11	0.16	0.23	-0.28	-0.10	-0.38
log Kow	0.30	0.31	0.19	0.06	0.09	0.35
nAcid	0.17	0.16	-0.46	-0.31	-0.32	-0.22
nAromBond	0.00	0.00	0.00	0.00	0.00	0.00
nB	0.00	0.00	0.00	0.00	0.00	0.00
nBr	-0.01	-0.11	0.13	-0.03	-0.01	-0.03
nC	-0.03	0.01	-0.01	0.02	0.16	0.21
nCl	-0.14	-0.10	-0.30	0.15	0.18	0.03
nF	0.39	0.42	0.37	-0.02	0.00	0.06
nH	-0.15	-0.13	-0.24	-0.04	0.03	0.30
nHBAcc	-0.08	-0.06	-0.22	-0.25	-0.05	-0.12
nHBDon	-0.10	-0.04	-0.04	0.06	0.13	-0.26
nI	-0.16	0.01	0.00	-0.18	-0.14	-0.26
nN	-0.14	-0.13	0.40	0.48	0.50	0.00
nO	0.12	0.13	0.01	-0.63	-0.47	-0.31
nP	-0.09	-0.13	0.00	-0.01	0.03	0.07
nS	0.15	0.21	0.04	0.02	0.06	-0.23
nX	0.00	0.10	0.04	0.06	0.13	-0.06
naAromAtom	0.00	0.00	0.00	0.00	0.00	0.00

Attachment VII Contribution of descriptors values in the prediction of the response factor

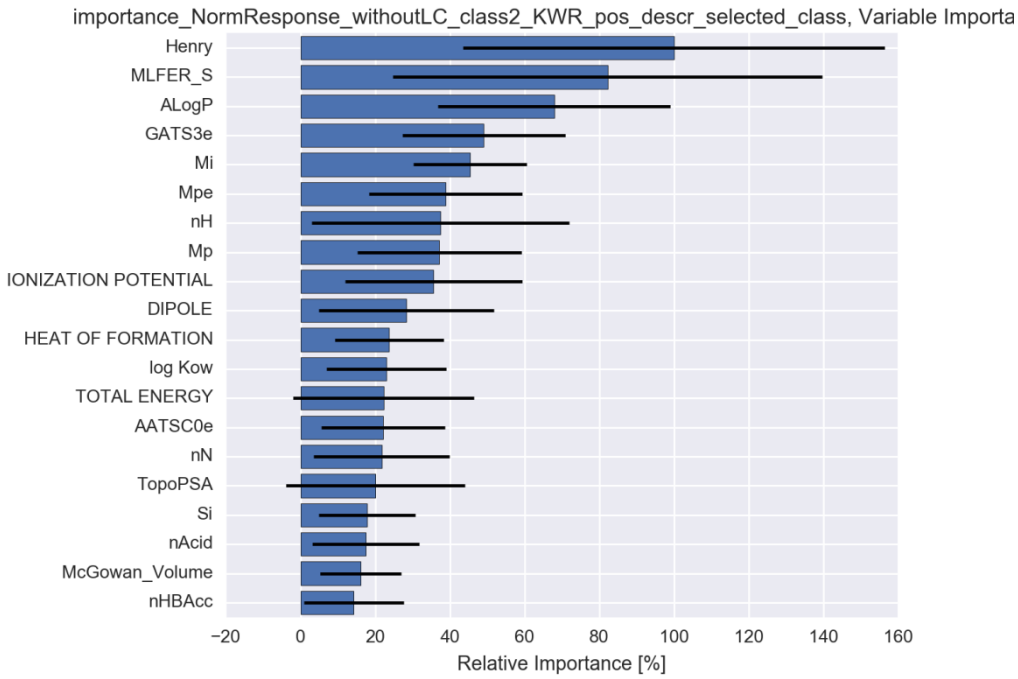
Positive ionisation mode with LC-column (2 classes)



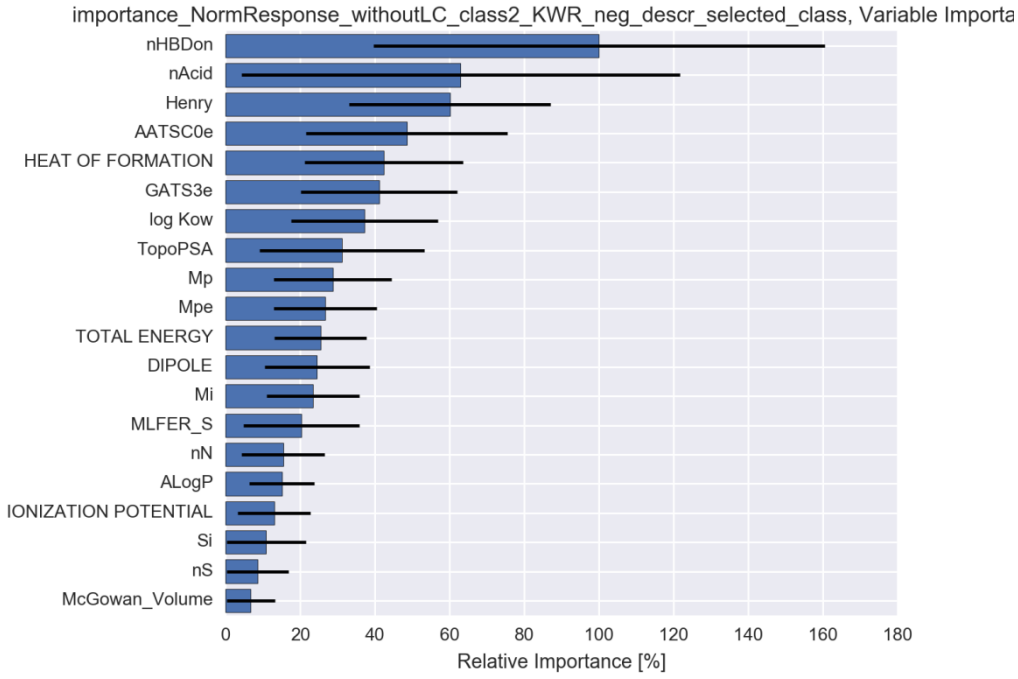
Negative ionisation mode with LC-column (2 classes)



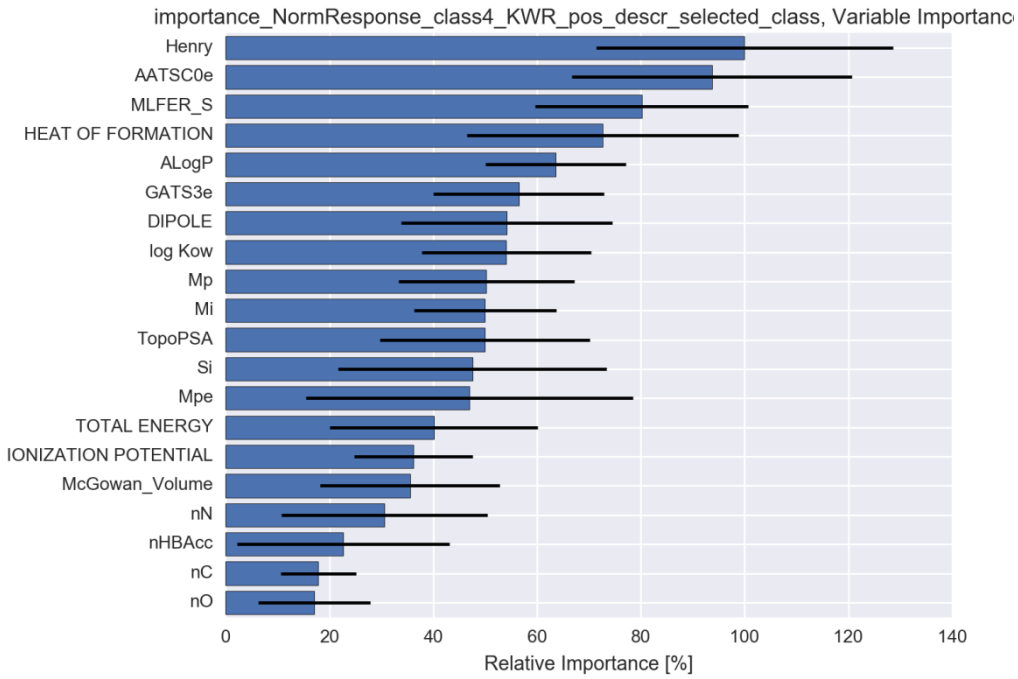
Positive ionisation mode, direct injection (2 classes)



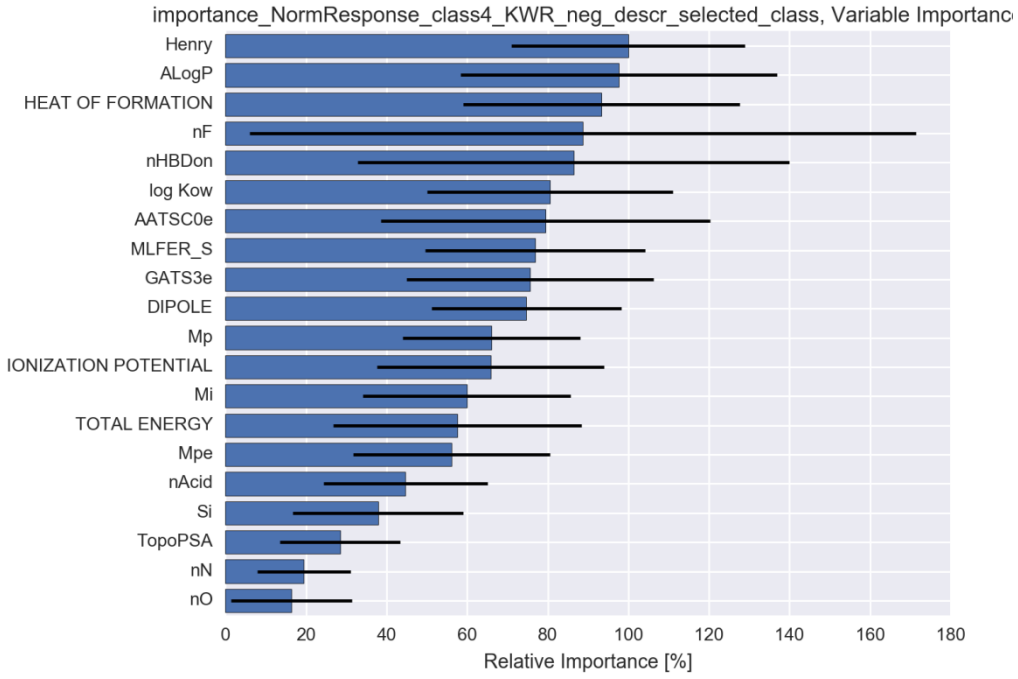
Negative ionisation mode, direct injection (2 classes)



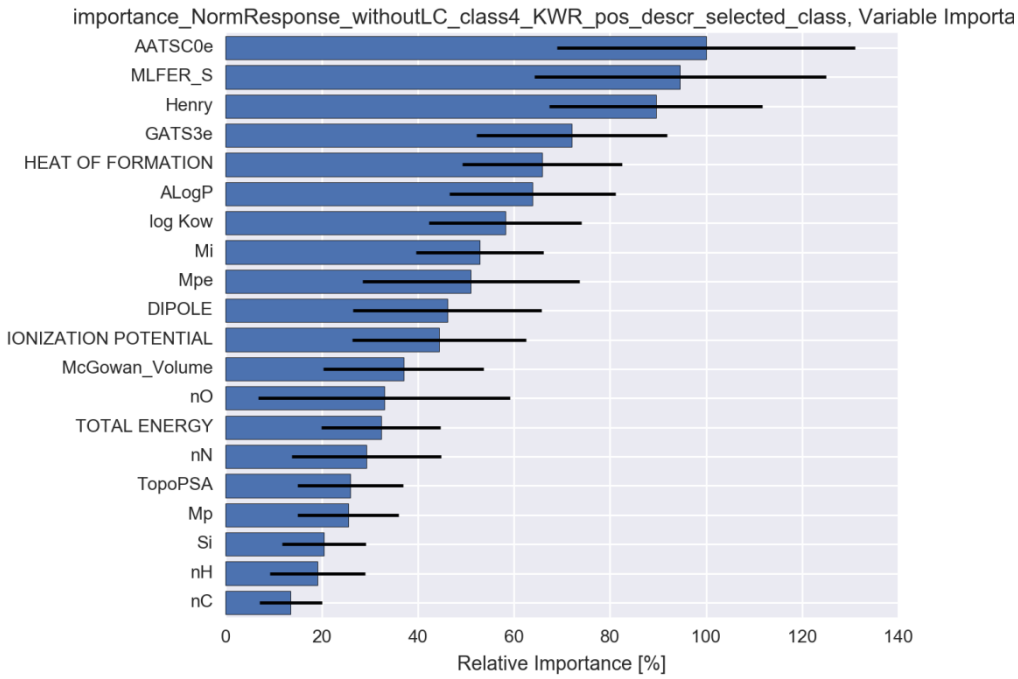
Positive ionisation mode with LC-column (4 classes)



Negative ionisation mode with LC-column (4 classes)



Positive ionisation mode, direct injection (4 classes).



Negative ionisation mode, direct injection (4 classes)