Department of Materials- and Environmental Chemistry
Unit of Analytical Chemistry

Widening the Horizon of Non-Targeted Environmental LC/ESI/HRMS Analysis: Preparation for the NORMAN Interlaboratory Comparison on Semi-Quantification

15 Credits
Bachelor's Thesis

Louise Malm

Supervisor: Anneli Kruve, Associate Professor
Stockholm, July 2020

Table of Contents

Abbreviations

DMSO	Dimethyl sulfoxide
ESI	Electrospray ionisation
HRMS	High resolution mass spectrometer
IE	Ionisation efficiency
ILIS	Isotope labelled internal standards
LC	Liquid chromatography
MeCN	Acetonitrile
MeOH	Methanol
RF	Response factor
RT	Retention time
SI	Supplementary information
SPE	Solid phase extraction
TPs	Transformation products

Abstract

The number of chemicals in daily use increases and surface water worldwide is polluted by an ever-widening spectrum of compounds, from pharmaceutical residues and beauty products to pesticides and artificial sweeteners. The quality of the surface water influences the purity of our drinking water and also puts higher demands on wastewater treatment plants. The NORMAN Network[1] has compiled a list of more than 65 000 compounds suspected to be found in surface water. Detecting all of these compounds requires various analytical techniques, of which liquid chromatography electrospray ionisation high resolution mass spectrometry (LC/ESI/HRMS) is the preferred technique for water analysis. However, ordinary targeted analysis and quantification with standard substances is neither time nor cost effective for so many analytes. Furthermore, such analysis cannot detect new, emerging pollutants or ones for which there are no standards available. Therefore, non-targeted analysis is being increasingly employed. Together with semi-quantification of the analytes of interest, it is possible to prioritise which analytes are most important to identify.[2,3] Though there are semi-quantification methods available for use, none are yet standardised, and therefore not ready for wide-community use.

The NORMAN network is arranging a large interlaboratory trial for semi-quantitative non-targeted environmental analysis in 2020 to 2021. The intention of this trial is to make semi-quantification results comparable between different instruments and laboratories, but also to validate the proposed semi-quantification methods. The current thesis is a preparation for the collaborative trial, with focus on selecting the compounds to be used in the large-scale comparison. Furthermore, four semi-quantification approaches will be tested and evaluated.

1. Introduction

Even though our planet is largely covered in water, access to freshwater in some areas is limited and the scarce quantities available is steadily contaminated. Water pollutants from an ever growing spectrum of contaminants is not only a threat to global health in the sense of increased shortage of drinking water and sanitation,[4] but it also threatens life in the waters as well as the animals that eat for instance contaminated fish.[5] Many contaminants, up to approximately 70 000 according to Schwarzenbach et al. (2006),[6] come from common household products used daily such as sunscreens, medicines, detergents, etc. via the sewage systems and insufficient treatment plants. Agriculture with associated pesticides and fertilisers, together with industry are, however, responsible for the majority of pollutants, releasing hundreds of millions of tons to the environment annually.[6] Many of these compounds escape through wastewater treatment plants and end up in the surface water. Additionally, many chemicals degrade, either naturally in the environment or during water purification,[7] and thereby give rise to a number of transformation products (TPs) with supposedly similar structure as their parent compound. A comprehensive list of suspected compounds found in water by the NORMAN Network[1] contains more than 65 000 entries, giving an indication of the seriousness of the problem. Many of these compounds are TPs, mainly from pesticides, though contaminants from other groups like pharmaceuticals most likely also undergo degradation.[8,9] Furthermore, TPs are sometimes both more toxic and more abundant than their parent compound,[10] thus they should not be ignored.

Some steps to control the amount of pollutants released to the aquatic environment are in place, e.g. the Water Framework Directive in EU and Clean Water Act in the US. Still, the abovementioned organisations only control 48 compounds[11] and 126 compounds,[12] respectively, i.e. merely a tiny fraction of all compounds currently contaminating the surface waters.[9] Moreover, TPs are seldom included in water protective plans, mainly due to unknown structure[8] or toxicity.[10]

One of the more common techniques to analyse water samples for contaminants is liquid chromatography electrospray ionisation high resolution mass spectrometry (LC/ESI/HRMS) due to the polar nature of many contaminants.[8] Furthermore, a higher diversity of analytes that can be detected in LC/MS compared to gas chromatography mass spectrometry (GC/MS), since the compounds do not have to be volatile.[13] To increase the contaminants controlled, including more TPs, it is not feasible to use ordinary targeted analysis, i.e. to only look for a couple of selected pollutants in the sample. Neither can this kind of analysis find new emerging compounds, since it only finds *what* it's looking for. Instead, non-targeted or suspect screening methods are used more and more. However, a crucial problem remains, namely quantification. It is important to know both *which* pollutants are found in water, as well as *how much* of them there are. A solution to this lies in semi-quantification, which unlike "real" quantification where analytical standards are used, estimates the amount of the analytes without standards. This is an ongoing research topic, where new approaches for semi-quantification are just being, or have very recently been, developed. In this study, four of those strategies will be evaluated, which are further described in section 1.5. below.

1.1. Target, Suspect and Non-Targeted Screening

Target screening is a method to look for predetermined compounds, and with the help of reference standard compounds it allows quantification of the analytes. This is the default method to screen for controlled contaminants in water; however, this approach is not sufficient to screen for all contaminants that can be found in surface water.

In order to find more, and above all new, contaminants in water, non-targeted screening needs to be employed. This procedure is much more time consuming than targeted screening, especially the data processing part.

For identification, the chromatographic peaks first need to be sorted and combined according to which species they belong to.[8] This is done by a software, usually from the same manufacturer as the hardware, e.g. Compound Discoverer (Thermo Fisher Scientific™, USA). However, free software can also be used to combine the chromatographic peaks, e.g. from https://www.rformassspectrometry.org/. This step gives a set of possible molecular formulae for each peak, and after additional filtering, e.g. matching isotope pattern, one molecular formula is assigned to each peak. Identification by molecular formula reaches confidence level 4 according to a scale proposed by Schymanski et al. (2014),[14] as seen is figure 1.

To obtain a tentative structure of the contaminant, further processing is needed, usually by analysing the fragmentation in MS^2 spectra. In the rare case that only one tentative structure is obtained from MS^2 experiments, and additional evidence support that structure, confidence level 2 might be achieved. If instead more than one possible structure is procured, the fragmentation spectra needs to be compared to reference spectra to reach level 2 identification. The level is, therefore, called probable structure, indicating that the structure is assigned with rather high confidence, but not as high as the confirmed structure level. To reach the highest confidence level, the probable structure needs to be confirmed with the reference standard.

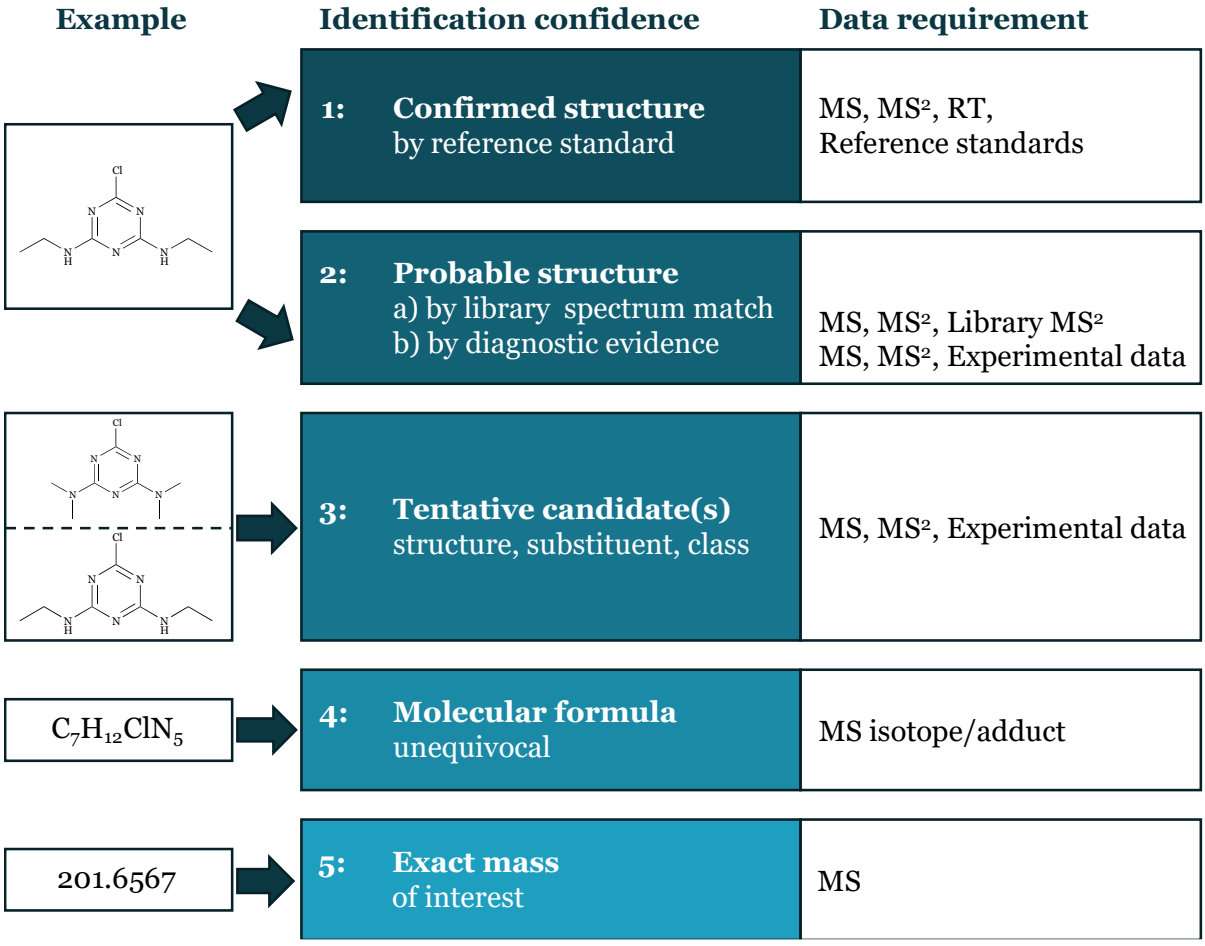


Figure 1. *The confidence levels for identification with their respective data requirements, as proposed by Schymanski et al.[14] As seen, a (tentative) structure is only available from level 3 and higher.*

The steps taken from level 3 to full identification usually requires searching databases for a match and is therefore not really "true" non-targeted analysis but rather suspect analysis. The

main difference between suspect- and non-targeted screening are that in suspect screening there is an idea of what to search for, e.g. from suspect lists or databases, whereas non-target analysis is performed with no prior information to what to search for.

Understandably, data treatment in non-targeted and suspect screening is tiresome, and some compounds may be overlooked or wrongfully dismissed in the process. Similarly, steps taken prior to data analysis, e.g. sampling, preparation, separation and detection, can also introduce errors.[8]

As some compounds, especially in water samples, are quite unstable it is generally important to analyse the sample as shortly after sampling as possible. If too long time passes between sampling and analysis, some compounds might degrade and thus cannot be detected. This is especially important for surface water, since this kind of water is fairly "new" as opposed to the much older groundwater. Furthermore, analysing in close proximity to sampling also minimise the risk for contamination.

Most often some kind of sample preparation is required to concentrate the sample, or many compounds would not be detected as their concentration is below instrument sensitivity.[10,15,16] Solid phase extraction (SPE) is a commonly used preconcentration technique, and is also used in this study. The choice of cartridge highly influences which compounds will be detected in the analysis and must therefore be carefully considered. However, since the analytes are unknown in non-targeted analysis, choosing the solid phase can be quite hard. To prevent the removal of analytes of interest, extracts from cartridges of different polarity can be combined. Preferably, a multilayer cartridge can be used, where the different layers have different polarity.[7] A standard cartridge used in non-targeted screening of water samples is OASIS® HBL, since it contains a copolymer with both hydrophilic and hydrophobic moieties.[17]

Another source to false negatives is introduced in the LC/HRMS process, or rather in the ionisation process. Since the mass spectrometer can only detect ions, compound that do not ionise well enough will not be detected, but this does not mean that the compound is not present in the sample. It is also important to consider the choice of column used in LC, to obtain the best chromatographic separation.

1.2. NORMAN Network

The NORMAN Association is a "network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances", i.e. a network of institutions that in some way deals with emerging substances. It is a non-profit organisation, aiming to collect and maintain information and data of new environmental substances, as well as promote further development of measurement and monitoring tools. The NORMAN Network manages or contributes to many databases, e.g. MassBank Europe (mass spectra of emerging substances) and Suspect List Exchange (various lists of substances for suspect screening and prioritisation), free for anyone to use. Furthermore, they seek to improve and exchange knowledge of analytical methods and approaches, mainly by interlaboratory studies. Collaborating laboratories are from all over Europe and North America.[18]

One of the interlaboratory collaborations arranged by the NORMAN Network this year (2020) is on semi-quantitative non-targeted analysis with LC/ESI/HRMS,[19] with the intention to make semi-quantification strategies comparable between different instruments and laboratories.

1.3. Interlaboratory Comparisons

There are different purposes for doing interlaboratory comparisons, which needs to be established before designing such studies. The goal can vary from determining the physical properties of an artefact to determining random or systematic variations in measurement results,[20] or to evaluate either a specific measurement method or a new reference material.[21]

In this particular case, the objective is to validate the semi-quantification approaches currently used and make the results comparable between laboratories. To do this, each participating laboratory will receive an aliquot from the same water sample, spiked with 67 chosen compounds. Of these, 37 will have a, for the laboratory unknown concentration, which will be determined using the quantification strategies.[22] The estimated concentrations will then be compared with the actual concentrations for each of the 37 compounds, to evaluate the semi-quantification approaches. Since the plan for the interlaboratory comparison is to send already spiked water samples to the laboratories, it is important to know whether the substances are stable in water matrix. Therefore, the stability will be tested and evaluated during this pre-study.

1.4. Response factor

The response factor (RF) of a compound is the ratio between the detected peak area and the concentration of the compound (see equation 1), with the unit [signal unit/M], where the signal is in arbitrary units.

$$RF = \frac{peak\ area}{concentration} \qquad (1)$$

The RF is highly influenced by how well the compound ionise; therefore, since not all compounds ionise to the same extent, the peak area is not proportional to the concentration. This is giving rise to problems regarding quantification and is by extension one of the reasons why semi-quantification methods are needed. Many aspects are influencing the efficiency with which a substance ionises, e.g. the properties of the substance and the eluent, matrix effects and which ionisation source is used.[3,23] Important properties regarding the compound itself is hydrophobicity and basicity/acidity, as seen in figure 2. Regarding the eluent, the amount of organic solvent is significant, as well as the pH and buffer type.[8] Generally, the ionisation efficiency (*IE*) increases with the amount of organic solvent, since this makes the droplets in the ionisation source dry faster.[3]

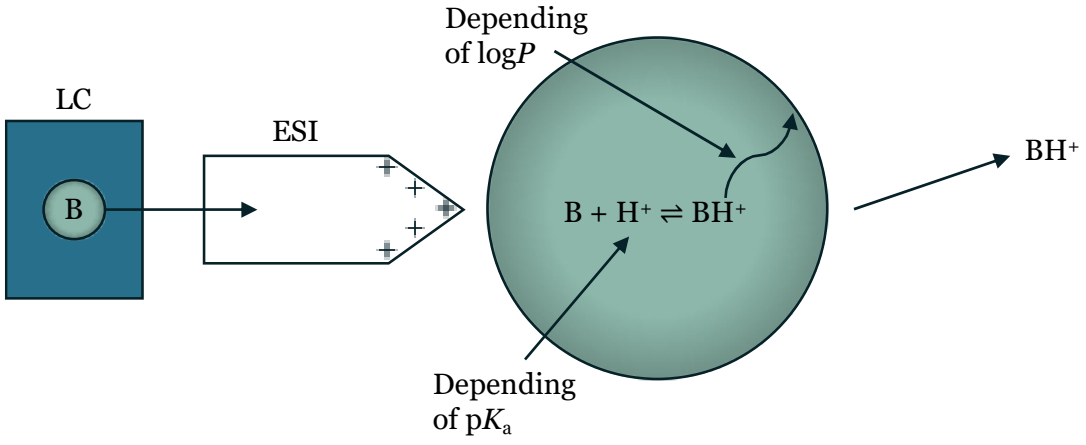


Figure 2. *Two important properties of the compound (B) that is influencing how well it will ionise. The basicity (pK_a) is influencing how well the substance will become protonated inside the droplet, and the hydrophobicity (logP) is influencing how easy the protonated ion will get to the surface of the droplet.*

1.5. Strategies for Semi-Quantification

In the interlaboratory trial, a total of six semi-quantification strategies will be tested and evaluated; however, in this smaller pre-study, only four of those will be examined due to the current technical availability. Namely, two of the proposed approaches are still under development. The used semi-quantification strategies are described in more detail below, with their respective assumptions and limitations displayed.

1.5.1. Structurally Similar Compounds

This approach takes advantage of the assumption that compounds with structural similarity have similar RF. This means, that the most structurally similar compound from the standards mix will be used to quantify the compound in question, according to equation 2:

$$c_{suspect\ compound} = \frac{peak\ area_{suspect\ compound}}{RF_{most\ similar\ standard\ compound}} \quad (2)$$

The most similar compound is found by comparing the functional groups of the suspect compound with all of the standard compounds, together with some other 2D chemical descriptors like the distance between functional groups.[22] An online tool[24] is available to search for the structurally most similar compound from the University of Athens standards database, down to a similarity of 5%. This tool requires the SMILES string of the suspected compound as an input. In this study, the list of similar compounds found by the tool is compared with the list of compounds in the standards mix to find the most structurally similar standard available. Additionally, the similarity score (%) and a maximum expected error is obtained.

A disadvantage with the online tool is that it calculates the similarity based on the *number* of similar functional groups but does not consider *which* groups are similar or dissimilar. E.g. the most similar compound may lack a very hydrophobic or basic group compared to the suspect compound, and still have a very high similarity score. This disadvantage might make this approach prone to errors.

1.5.2. Parent Compound – TPs

This strategy is quite similar to the previous but is instead taking advantage of the fact that TPs originate from a parent compound. The parent compound is then used, similarly as the structurally similar standard, to estimate the concentration of its TPs in the unknown mixture. This approach presumes that TPs are structurally similar to their parent, and thus are assumed to behave similarly. With this assumption, the RF of the parent is directly applicable for the TP, giving the following equation to calculate the concentration of the transformation product:

$$c_{TP} = \frac{peak\ area_{TP}}{RF_{parent\ compound}} \quad (3)$$

Dahal et al. (2011) has previously applied this strategy on drugs and their metabolites, with up to a 4-fold error.[25] However, this approach cannot be used as a single semi-quantification method in true non-targeted screening, as it is solely applicable for compounds that are the result of degradation.

1.5.3. Close Eluting Compounds

The third semi-quantification approach is based on previous work by Pieke et al. (2017),[2] and uses the response factor of the standard with the closest RT to the suspect compound, as per equation 4:

$$c_{suspect\ compound} = \frac{peak\ area_{suspect\ compound}}{RF_{closest\ eluting\ standard\ compound}} \quad (4)$$

Here, the estimated concentration of the suspect compound is not necessarily based on the standard with the most similar structure, but rather on the standard with the most similar chromatographic properties. The assumption is that compounds with similar retention time have resembling *IE* as well, and consequently, the RF of close eluting standard can be applied to the suspect compound. This approach has previously been employed with satisfying results, with a 2-fold average error.[2] Furthermore, this approach can be applied for all compounds detected in a sample, and even for compounds without a tentative structure.[22]

1.5.4. Predicting the Ionisation Efficiency

Lastly, a somewhat different approach will be tested, where the response factor of the suspect compound is predicted. The prediction is based on a model recently developed by Liigand et al. (2020),[23] which uses 2D PaDEL descriptors of the compound, eluent descriptors and random forest regression. Especially important parameters regarding the eluent are pH, eventual presence of NH_4^+, as well as viscosity, surface tension and polarity index. The *IE* can be calculated in R (© The R Foundation) based on the previously developed model, or by an online tool provided by Quantem Analytics.[26] The known concentrations of the standards will be used to calibrate the algorithm used for predictions to match the analytical method used.

Similar to some of the abovementioned strategies, this strategy can be used in suspect screening, i.e. when a tentative structure is available. However, a drawback is that currently, the algorithms can only predict the response factor for protonated ions. This means that this approach is not applicable for compounds that form only other major adducts, e.g. sodium or ammonium adducts.

2. Materials and Methods

2.1. Chemicals

The standard mix consisted of Guanylurea, Amitrole, Histamine, Chlormequat, Methamidophos, Vancomycin, Trichlorfon, Butocarboxim, Dichlorvos, Tylosin, Rifaximin, Spinosyn A, Emamectin B1a, Nigericin, Ivermectin B1a, TCMTB chosen by the University of Athens, and Atrazine, Octocrylene, Clarithromycin, Aspartame, Simvastatin, Sucralose, Irgarol, Caffeine, Carbamazepine, Benzotriazole, Metolachlor and Imazalil selected at Stockholm University.

The suspect mix consisted of Atrazine-desethyl, Atrazine-desethyl-2-hydroxy, Atrazine-desisopropyl, Atrazine-desethyl-desisopropyl, Atrazine-desethyl-desisopropyl-2-hydroxy, Atrazine-desisopropyl-2-hydroxy, Atrazine-2-hydroxy, 2-(Methylthio)benzothiazole, Progesterone, Butylamine, Haloperidol, Reserpine, Phenazine, Clotrimazole, Simazine, Efavirenz, Adenosine, Climbazole, Melamine, Metazachlor, Chlorothiazide, Metformin, 2-Methylbenzothiazole, Benzothiazole, Chlorpyrifos, 5-Methyl-1H-benzotriazole, 10,11-Dihydro-10-hydroxycarbamazepine, Sudan I, Ketoconazole, 5-Chlorobenzotriazole, Benzotriazole-5-carboxylic acid, Carbamazepine-10,11-epoxide, Metolachlor-ESA, Metolachlor-OXA, Omethoate, 2-Hydroxybenzothiazole and 2-Aminobenzothiazole.

The isotope labelled internal standard (ILIS) mix consisted of Atrazine-d_5, Caffeine-$^{13}C_3$ and Imazalil-d_5.

TCMTB was bought from Honeywell Fluka™, all other chemicals were from Sigma-Aldrich. All chemicals were of analytical standard.

Each chemical was dissolved in either Acetonitrile (MeCN), Methanol (MeOH), MilliQ water, Dimethyl sulfoxide (DMSO), 0.1% formic acid or 0.1M Hydrochloric acid (HCl), or in a mix of two of the mentioned solvents. The exact solvent for each compound can be found in table S1 in supplementary information (SI). All solvents were from Honeywell Riedel-de Haën™, except HCl that were from VWR Chemicals and ultrapure water which was purified with Milli-Q IQ 7000 (Merck, Darmstadt, Germany).

2.2. Samples

Solutions of all chemicals were prepared to a concentration of around 1000 ppm, mixed according to the mixes in section 2.1 and were lastly diluted. Final concentration for each compound and mix can be found in tables S1 and S2 in SI.

Water sample from Laduviken was obtained on June 16th, filtered twice using Munktell Filter Paper (Ahlstrom Munksjö) and stored at +4°C prior to sample preparation. Tap water samples were obtained from a tap at Stockholm University directly before extraction.

SPE was performed to concentrate the water samples, based on the SPE method used by Rousis et al. (2017),[27] using OASIS® HBL 6cc/150 mg cartridges (Waters corp., Milford, MA, USA). Cartridges were conditioned with 12.5 mL of MeOH followed by drying under ~ 0.2 bar vacuum for 10 min. The cartridge was equilibrated with 7.5 mL MilliQ water before 125 mL of sample water (Laduviken or tap) was loaded and passed with a flow rate of approximately 2 mL/min. Immediately after the sample, the cartridge was washed with 5% v/v MeOH in MilliQ water, and then dried under ~ 0.2 bar vacuum for 10 min. The extract was eluted with 7.5 mL of MeOH, and then evaporated under a gentle air flow at 30°C. Both extracts were later reconstituted in 3 mL MilliQ water, and the water samples (SB1 – SB4) were spiked with abovementioned mixes (standard, unknown, and ILIS mix). Two different concentrations for unknowns were used. Final concentrations can be found in table S2 in SI.

2.3. Instrumental

Spiked water samples were analysed with Dionex UltiMate™ 3000 UHPLC system consisting of an RS Pump, RS Autosampler and RS Column Compartment (Thermo Fisher Scientific™, USA). The chromatographic separation was carried out on a Kinetex 2.6 μm, EVO C18, 100 Å, 150 × 3.0 mm column (Phenomenex®, Torrence, CA, USA). The mobile phases used was A: 0.1 % formic acid in MilliQ water with pH 2.7, and B: MeCN. Eluent gradient started with 5 % B, and then gradually increased to 100 % B over 20 minutes, was held at 100% B for 5 minutes and then decreased back to starting composition over 0.1 min. Between each injection, the column was equilibrated for 5 minutes. Sampler temperature was 15 °C, column oven temperature 40 °C, flow rate was 0.35 mL/min and the injection volume were 5 μL.

Samples were measured in positive ESI mode on a Q Exactive Orbitrap HRMS (Thermo Fisher Scientific™, USA), in two scan ranges; *m/z* 60.0000 – 900.0000 and *m/z* 100.0000 – 1 500.0000, with a mass resolution of 120 000. In the ion source, spray voltage was 3 500 V, capillary and probe heater temperature were 320 °C, max spray current was 100 μA and the S-lens RF level was 50 %. The gas parameters were sheath gas: 35, aux gas: 3, and spare gas: 0 (all in arbitrary units).

2.4. Data Treatment

Peaks from the chromatograms were integrated using the processing tool in XCalibur (Thermo Fisher Scientific™, USA), based on the *m/z* for each compound, with a mass tolerance of 10 ppm. From the processed files, information of the ions detected ($[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, $[M]^{2+}$) and their respective peak area and retention time were exported to Excel (Microsoft corp., USA) and saved as .csv files for further analysis in R (© The R Foundation).

The detected compounds were semi-quantified in R Studio, according to the principles described in section 1.5. If the predicted concentration ($c_{\text{predicted}}$) was larger than the real concentration (c_{real}), then

$$prediction\ error = \frac{c_{predicted}}{c_{real}}. \quad (5)$$

If instead the real concentration was larger than the predicted concentration, then

$$prediction\ error = \frac{c_{real}}{c_{predicted}}. \quad (6)$$

The code for the semi-quantification methods is available in supplementary information.

3. Results

A total of 67 compound for standards mix and suspect mix were selected from the NORMAN suspect list, to evaluate four semi-quantification strategies used in suspect screening. Selection was based partly on the chemicals' probability to end up in the surface water, and on their relevance in daily life, e.g. pharmaceuticals, pesticides, cosmetics, etc. There was also a desire to cover as wide chemical space as possible, especially in the aspect of chromatographic and ionisation properties, i.e. hydrophobicity and basicity. Thus, selection was also highly influenced by the compounds' log*P* and pK_a. Due to technical reasons, two of the compounds, Cefoperazone and Avermectin B1a, were not available at the time for this thesis but will be included in the interlaboratory comparison. Before making the final decision of which compounds to include, test runs were performed at different concentrations to make sure that the substances were visible in the chromatogram. Three compounds, one from the standard mix and two from the suspect mix, were included even though they were not visible in the test runs. The standard compound, Chlormequat, is needed for one of the approaches in the interlaboratory trial. The two suspect compounds, Metolachlor-ESA and Metolachlor OXA was included as they were two of few TPs, that might be visible for other laboratories in the upcoming comparison. These three compounds are, however, excluded from the further discussion. Three additional compounds, that were seen in the test runs but not in the real runs, are also excluded from the discussion. These were Octocrylene, Butylamine and Melamine. The test runs also ensured that the concentrations chosen to spike the water samples with were in the linear range. This is very important for the RF to be applicable in the semi-quantification methods used.

Additionally, three isotope labelled standards (ILIS mix) were added to the samples to adjust for any variation in injection volume. Spiked water samples (SB1 – SB4) were analysed and the 37 compounds in suspect mix were semi-quantified according to the proposed approaches.

As can be seen in figure 3, the standard as well as the suspect compounds are well distributed, with both retention time and response factor for both mixes dispersed over a wide range.

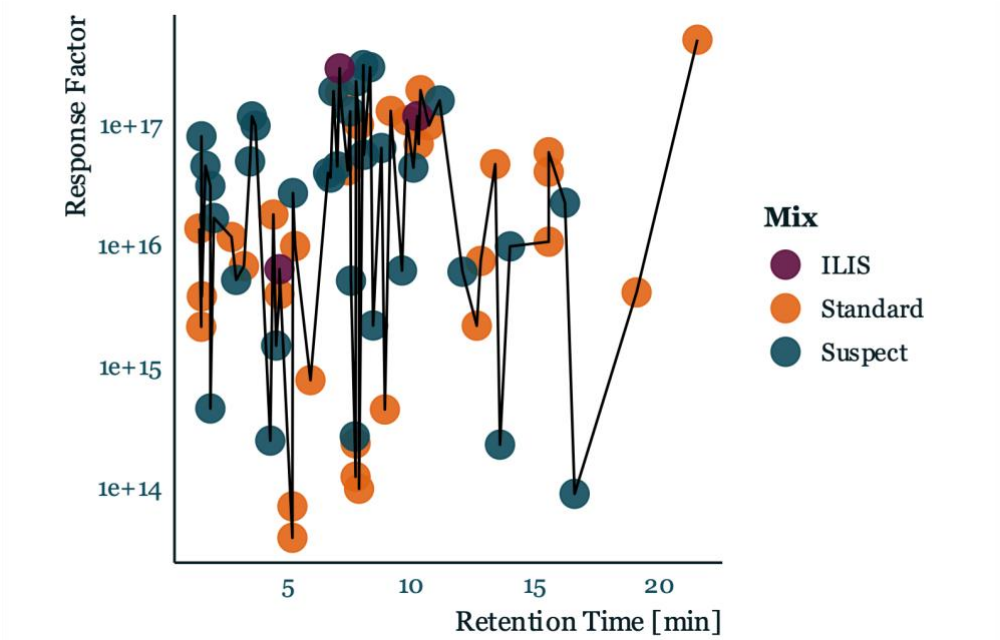


Figure 3. *Plot displaying the distribution of suspect- and standard compounds over RT and RF for sample SB1. Distribution for the other samples did not deviate significantly.*

Figure 4 displays how well the predicted concentrations correlate to the real concentrations, and in figure 5, a boxplot with the prediction errors for each approach is presented. The boxplot shows that the parent- and similar compound approaches give the lowest median error, while the close eluting standard approach appears to give the largest error. At the same time, all approaches give some unacceptable predictions, with over 1000-fold error for the close eluting standard and similar compound approach. The data in figure 4 are from one sample, however, data from the other three samples give similar correlation.

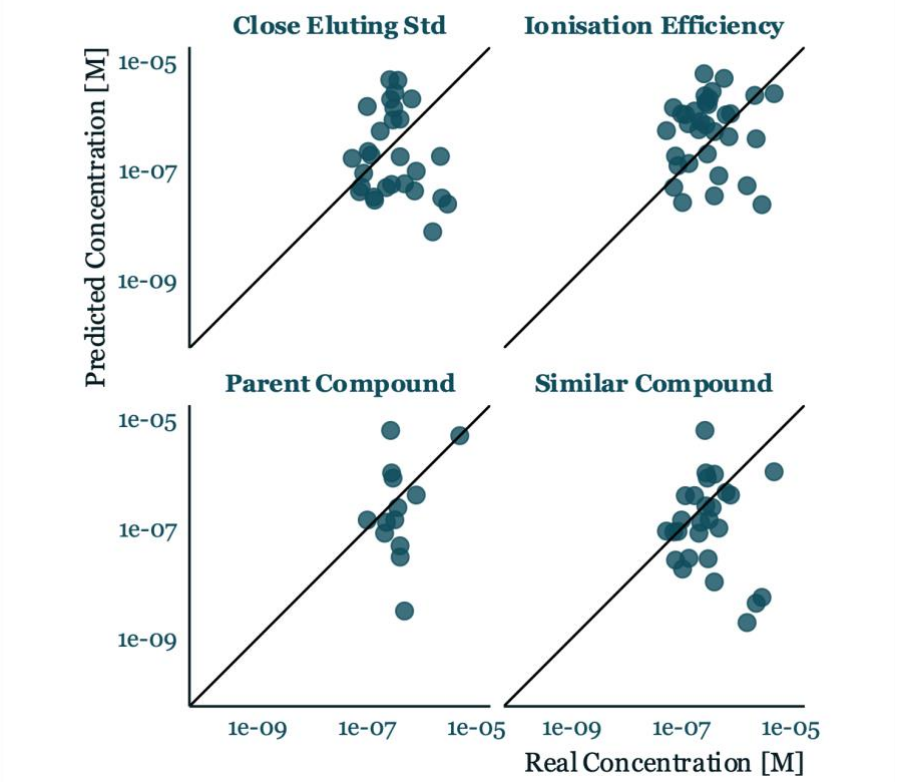


Figure 4. *Plot of the predicted concentration vs the real concentration for each semi-quantification strategy used, for sample SB1. The data from the other samples were very similar to this data.*

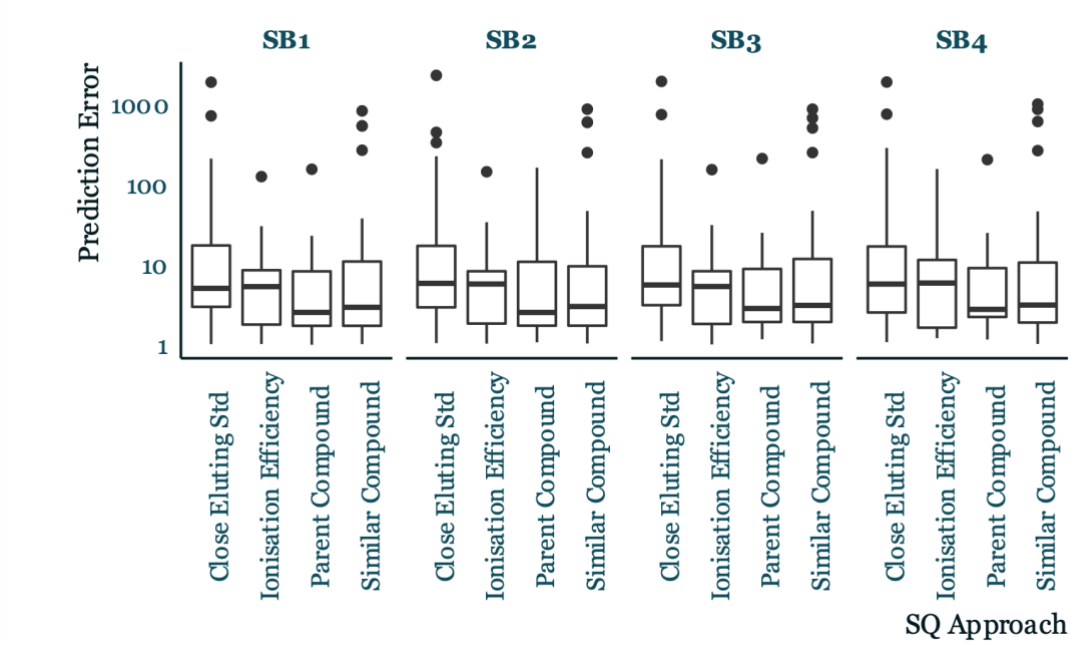


Figure 5. *Boxplot displaying the prediction error for each approach and each sample.*

Analysis of the samples was performed once a week for four weeks, to evaluate the stability of the compounds in water matrix. Stability testing is important for many reasons, in this case mainly to determine the suitability to send spiked water samples to the laboratories participating in the NORMAN collaborative trial. The stability is presented in figure 6, with two different approaches to normalise the peak areas.

Measurements from week 1 was ignored as all the intensities was much lower than the following weeks. Plots including this measurement can be found in SI, figure S1.

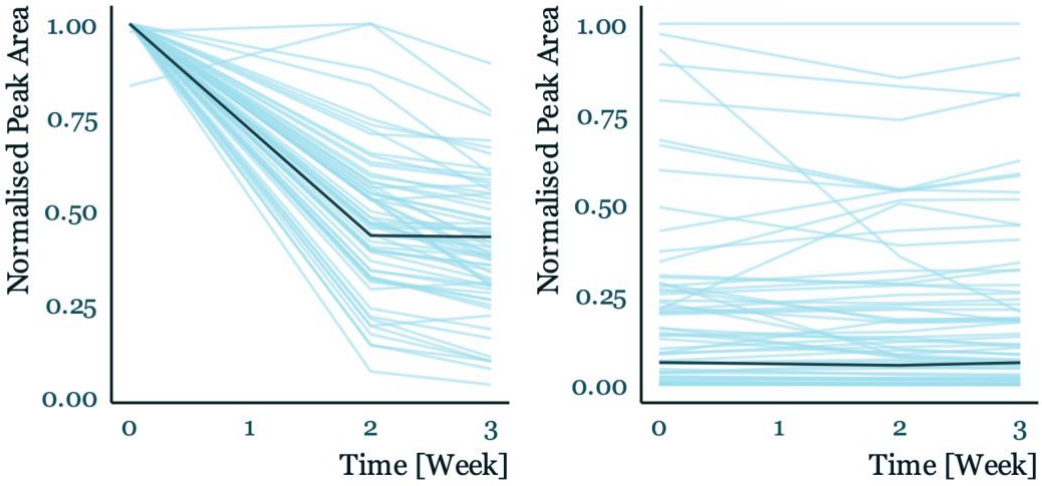


Figure 6. *Stability plots showing how the peak area for each compound changes over time. To the left the measurement with the largest peak area for each compound is used to normalise that compounds peak areas for the other measurements. Almost all compounds that are used for the normalisation is from week zero, with two exceptions: Benzothiazole and 2-Methylbenzothiazole. This normalisation approach assumes that the instrument is stable over the whole analysis period. The right plot has instead normalised the peak areas relative to the compound with the largest peak area (Metformin), with the assumption that this substance is stable.*

As seen, the different ways to normalise the peak areas give quite contradicting results. In the left-side plot in figure 6, almost all compounds appear to degrade quite rapidly, while in the right-side plot most compounds seem to be stable. The inconclusive results show that the stability needs to be further investigated during the interlaboratory trial with proper quantification approach.

4. Discussion

4.1. Compound Selection and Distribution

One of the main focuses for this study, and a very important aspect for the interlaboratory collaboration, was to select which analytes to include. As mentioned, the aim was to choose compounds with large variations in both chromatographic properties and relevance in the environment. The selection was also influenced by the semi-quantification methods that were to be used, i.e. parent – TP pairs were needed, as well as a good distribution between the standard and suspect compounds retention times. 18 of the standard compounds were already preselected by University of Athens; 2 of these were the ones that were not available at the time for the analysis. These 18 compounds are needed for calibration in one of the semi-quantification approaches suggested by their lab. Unfortunately, it was not possible to include that approach in this pre-study, as it has not yet been completely developed.

In figure 7, RF is plotted against RF for the standards and suspect analytes, to display the distribution of the two mixes. It can be seen that the majority of the analytes have quite high sensitivity. It would be desirable if more substances, especially from the suspect mix, would have had lower sensitivity. On the other hand, low responding compounds are generally harder to predict, and can also be harder to detect in the analysis, depending on the instrument used.

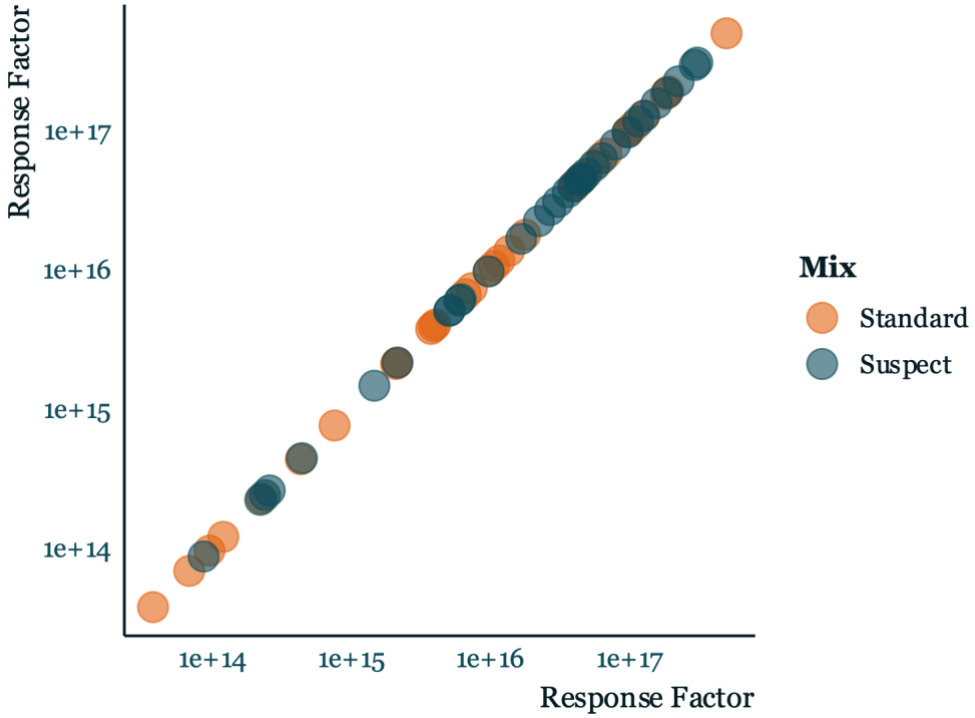


Figure 7. *Plot of the RF vs RF for the suspect and standard compounds, showing the distribution of the substances with respect to their sensitivity.*

Apart from wanting more analytes with lower response factor, the distribution between standards and suspect compounds seems to be quite satisfactory, based on the distribution of RF values. However, from figure 3, it can be deduced that some more late eluting suspect compounds would have been beneficial. After 12 min, 6 standards elute and 5 suspect compounds. The three data points at RT ≈ 15 min actually belong to the same compound, just different adducts.

There are no obvious compounds that could be interchanged between the suspect and standard mix. The two latest eluting substances seen in figure 3 (Ivermectin B1a and Nigericin)

both belong to the calibration mix from University of Athens and thus have to be in the standard mix. Similarly, the two lowest responding peaks in figure 7, are $[M+Na]^+$ and $[M+NH_4]^+$ adducts of Sucralose, so it would not be useful to substitute them for a suspect compound.

4.2. Prediction Errors

In figure 5, the boxplot displaying the prediction error for each semi-quantification approach, it would appear that best predictions are made with the parent compound approach. At least this approach is displaying the lowest median prediction errors across all the samples; however, this is not supported by the maximum errors or the mean errors, see table 1 . By these criteria, the ionisation efficiency approach appears to be most suitable, even though the median error is the second worst of the four approaches. The closest eluting standard approach exhibits the poorest results, with up to over 2000-fold prediction error for Carbamazepine-10,11-epoxide. This is supported by both table 1 and figure 5.

Table 1. *Mean-, median-, and max error for each sample and semi-quantification strategy.*

Approach	Mean error	Median error	Max error	Sample
Close Eluting Std	101.6	7.4	1880.8	SB1
Close Eluting Std	113.5	6.8	2286.4	SB2
Close Eluting Std	101.7	7.4	1924.3	SB3
Close Eluting Std	104.0	7.4	1886.6	SB4
Ionisation Efficiency	11.1	6.1	125.3	SB1
Ionisation Efficiency	12.2	6.1	143.9	SB2
Ionisation Efficiency	11.8	5.5	153.0	SB3
Ionisation Efficiency	12.7	6.5	155.8	SB4
Parent Compound	34.3	2.8	184.2	SB1
Parent Compound	35.1	2.7	178.5	SB2
Parent Compound	36.1	3.0	210.2	SB3
Parent Compound	31.3	2.8	203.9	SB4
Similar Compound	84.1	3.4	825.2	SB1
Similar Compound	89.2	3.1	867.3	SB2
Similar Compound	87.6	3.3	868.1	SB3
Similar Compound	100.1	3.3	1004.1	SB4

4.2.1. Correlation of the Errors with Chromatographic Properties

Analysis of the data revealed that there is insignificant correlation between the retention time and the prediction error, as seen in figure 8. The RT for the analytes with the highest errors ranged from 1.85 min for Atrazine-desethyl-desisopropyl-2-hydroxy to 16.58 min for Chlorpyrifos.

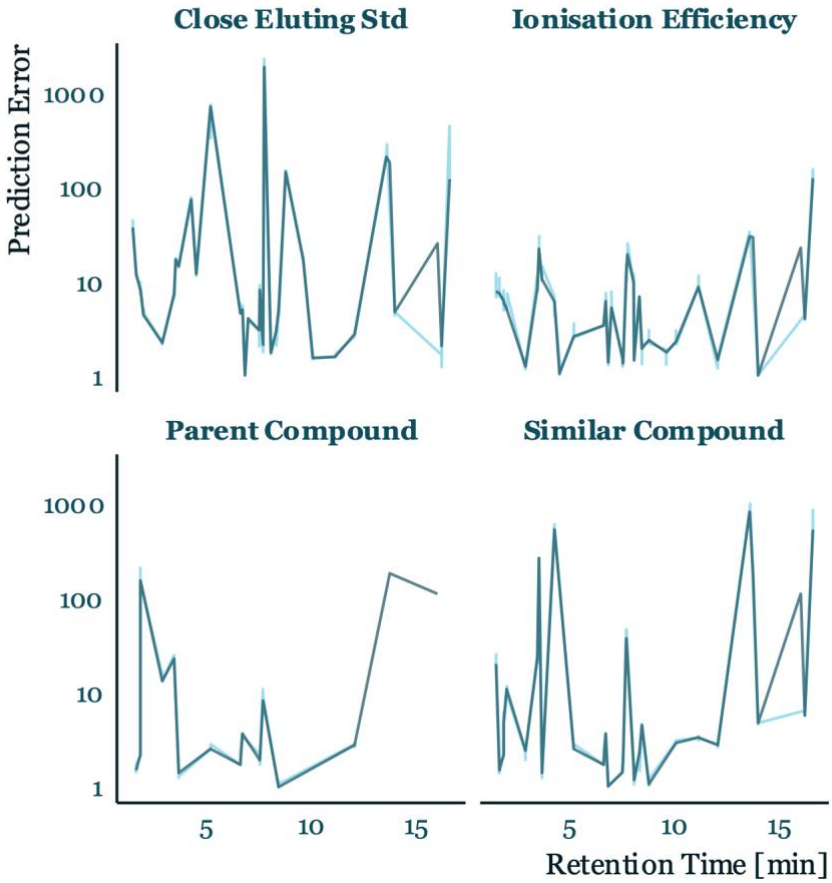


Figure 8. *Plot showing that there is no trend to be seen with regards to the prediction error and the retention time. As seen, the pattern is basically the same across all samples.*

All of the compounds yielding largest prediction errors are quite weak bases, with pK_a values from -4.2 to 4.6. This weak basicity makes the compounds harder to ionise, thus they yield a lower response factor, which in turn makes them harder to predict. This has been shown previously by both Liigand et al. (2020),[23] and by Kruve & Kaupmees (2017) for the ionisation efficiency prediction models.[28] Though this connection has been shown for *IE* prediction models only, it appears to be the case for the other approaches too, as seen in figure 9. The trend is therefore most distinct in the *IE* approach, but is also seen in the other approaches. However, some of the substances with unacceptably high error still has high response factor, which is especially clear in the closest eluting standard and the similar compound approaches. In fact, three out of the five substances with highest error in the closest eluting standard approach are in the high RF range. These are, ordered based on RF, Atrazine-desisopropyl (mean RF ≈ 3e+16, mean error ≈ 630-fold), Simazine (mean RF ≈ 6e+16, mean error ≈ 145-fold) and Carbamazepine-10,11-epoxide (mean RF ≈ 2e+17, mean error ≈ 2000-fold). In the similar compound approach, only one of the top five compounds with highest error also had high sensitivity, namely Omethoate (mean RF ≈ 1e+17, mean error ≈ 260-fold). The trend that the error decreases as the response factor increases, seems correct up to a certain value of RF, where the error starts increasing again. This is seen for all strategies, including the ionisation

efficiency, and thereby indicates that the prediction error cannot be explained by response factor alone.

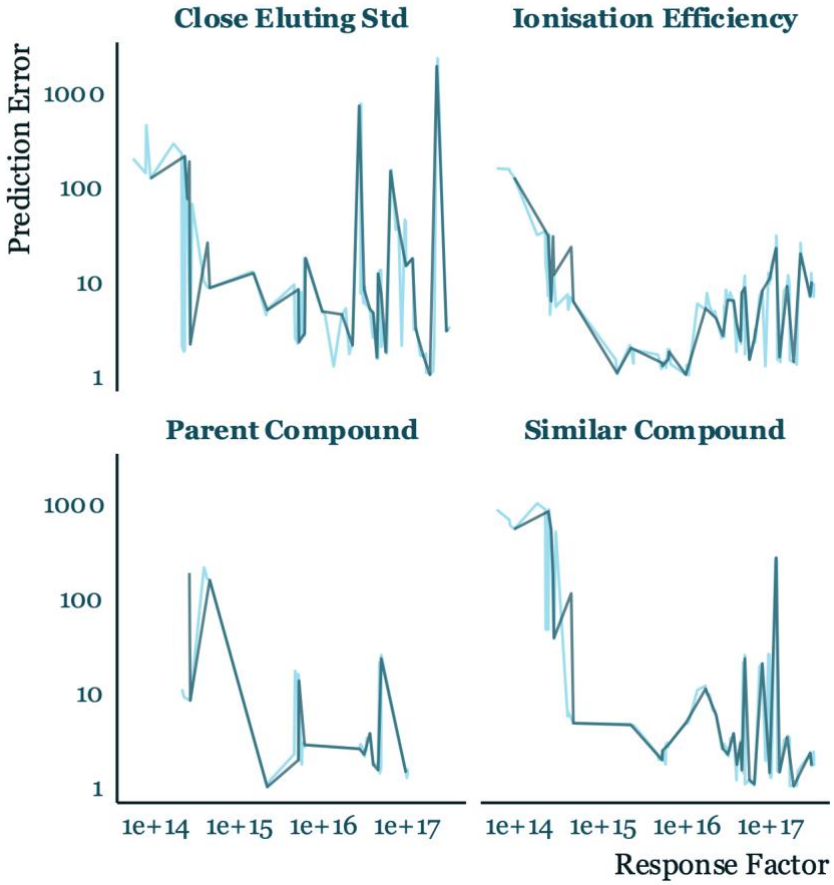


Figure 9. *Plot displaying an overall trend that analytes with low sensitivity gives high prediction errors. This trend is similar for all four samples and approaches.*

4.2.2. Adduct Formation and Semi-Quantification Error

Some of the substances with the highest errors formed adducts, e.g. Atrazine-2-hydroxy formed sodium adduct, Efavirenz, Chlorothiazide and 2-Hydroxybenzothiazole formed ammonium adducts, and Carbamazepine-10,11-epoxide and Omethoate formed both sodium and ammonium adducts. Of these, only the Chlorothiazide ammonium adduct had higher intensity than the protonated ion, which was the peak that was integrated and used in the semi-quantification. Therefore, this might explain high prediction error for Chlorothiazide. The Efavirenz ammonium ion was around 20% of the intensity of $[M+H]^+$ ion, whereas the adducts of the other analytes were less than 10%. Thus, they were deemed as not significant to explain the high prediction errors. The mass spectra of the compounds with highest error was also investigated, to determine whether any insource fragmentation occurred. Though there were some compounds that displayed fragmentation pattern, none of them were significant, i.e. less than 10 % intensity.

4.3. Structural Similarity

In table S4 in supplementary information, the suspect compounds are listed, together with the closest eluting/parent/most similar standard compound. In table S5, structures of all compounds that were visible in the analysis is displayed. With the exception of the parent compound approach, many of the compounds with the highest error across the semi-quantification methods are the same. E.g. Efavirenz and Chlorpyrifos, with prediction error

ranging from approximately 30- to 100-fold and 120- to 670-fold, respectively, are among the five highest substances with highest error for all approaches except the parent compound one. It should be mentioned that another compound with the same thiophosphate group as in Chlorpyrifos, previously has shown similar poor prediction results. In unpublished work by Kiefer et al.,[29] a similar compound, Diazinon, was the one with the highest error for both parent compound approach and ionisation efficiency predictions. This leads to suspicions that something with the thiophosphate moiety makes molecules harder to semi-quantify potentially referring to uninvestigated aspects in the ionisation mechanism. It might also explain why Omethoate is among the compounds with highest error, as is possess a similar phosphate moiety.

A qualitative analysis was performed for the five substances giving the highest error, to evaluate their respective similarity. For the parent compound approach, this analysis was done for three compounds only, since there were only three that gave error that were deemed unacceptable, i.e. more than 10-fold. In all, the closest eluting standard, *IE*, and similar structure approaches had 12, 6-11, and 7-8 compounds respectively that gave error higher than acceptable.

4.3.1. Parent Compound Approach

For the parent compound approach, the TPs look fairly similar to their parent compound, as seen in table 2. Yet, according to the similarity online tool described in section 1.5.1.,[24] the Atrazine TPs both shared more similarity with Guanylurea than with Atrazine. Qualitatively, significant differences in the functional groups can be observed as well for the TPs in table 1. Both the isopropyl- and the ethyl group has been cleaved in Atrazine-desethyl-desisopropyl as well as for Atrazine-desethyl-desisopropyl-2-hydroxy. For the latter TP, the chlorine atom has also been exchanged for a hydroxy group. Similarly, as TCMTB degrades to 2-Aminobenzothiazole, it loses quite a large moiety, which might explain why this TP gives such a high prediction error.

Comparing the hydrophobicity, it is seen that it decreases for all of the TPs compared to the parent compounds. However, this is more pronounced for the Atrazine TPs than for the transformation product of TCMTB. From the hydrophobicity aspect, Atrazine-desethyl-desisopropyl-2-hydroxy indeed seems to be more similar to Guanylurea than Atrazine. On the contrary, comparison of the basicity shows quite the opposite. Even though the pK_a values differs between parent and TP, they are all in the range of weak bases.

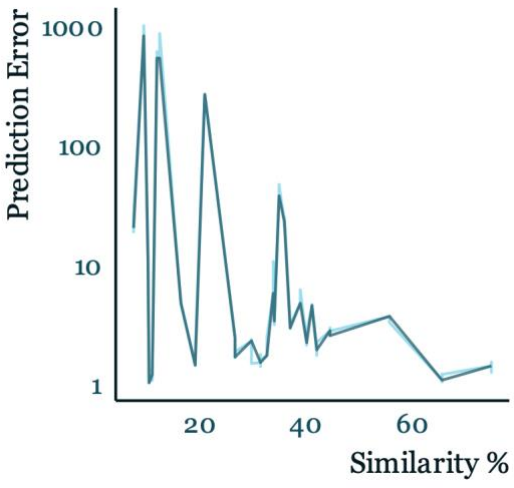


***Figure 10**. Plot showing the correlation between similarity score and prediction error. The trend appears to be that the prediction error decreases as the similarity score increases.*

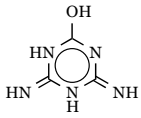
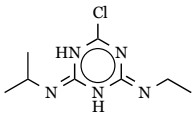
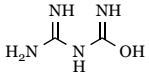
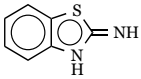
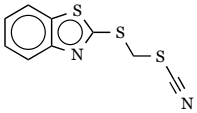
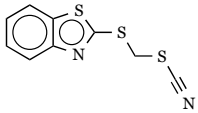
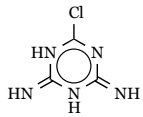
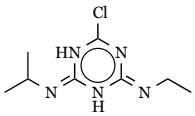
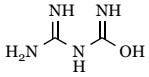
4.3.2. Similar Compound Approach

As seen in table 3, none of the most similar structures correlates well with the structure it is supposedly most similar to, at least not for the five analytes with the highest prediction error. Basically, the suspect compound shares some aromaticity with its most similar standard compound, and some common atoms can be found in the structures. The similarity score is ranging from 9.5 % (for Efavirenz and Imazalil) to 35 % (for 2-Hydroxybenzothiazole and Benzotriazole), which is not such a high score. The lower the similarity score is, the higher the maximum expected error becomes. Still, none of the suspects

are within the maximum expected error range. In fact, the mean error for Efavirenz is approximately 50x higher than the maximum expected error, and for 2-Hydroxybenzothiazole the mean error is approximately 5x higher than expected.

Based on the results in this study, it seems to be a correlation between the similarity score and the error, as seen in figure 10, which is hardly surprising. However, there are exceptions, e.g. Metformin, which is 7.6 % similar to its most similar compound Caffeine. Still, the mean prediction error is approximately 20-fold, i.e. much less than for Efavirenz although the similarity score is lower. This is likely to result from probability in obtaining similar response factors for two well ionising compounds, not from significant structural similarities.

Table 2. *Structural, hydrophobicity and basicity comparison of the transformation product and the parent compound, with average error over 10-fold. The most similar compound is also included to make the discussion easier to follow. pK_a- and logP values were all calculated in ChemAxon.*

TP	Parent	Most similar std
Atrazine-desethyl-desisopropyl-2-hydroxy	Atrazine	Guanylurea
		
Mean error: 182.6		
log*P*: -4.5	log*P*: 2.2	log*P*: -2.4
pK_a: 3.1	pK_a: 4.2	pK_a: 9.8
2-Aminobenzothiazole	TCMTB	TCMTB
		
Mean error: 23.4		
log*P*: 1.9	log*P*: 3.7	log*P*: 3.7
pK_a: 4.48	pK_a: 1.0	pK_a: 1.0
Atrazine-desethyl-desisopropyl	Atrazine	Guanylurea
		
Mean error: 14.8		
log*P*: -0.25	log*P*: 2.2	log*P*: -2.4
pK_a: 4.58	pK_a: 4.2	pK_a: 9.8

4.3.3. Closest Eluting Standard Approach

The similarity between the suspect compound and the closest eluting standard for this approach is the worst across all the approaches. In table 4, it is seen that there is practically no similarity at all between the pairs. However, this strategy is not based on finding compounds with structural similarities, but rather to find compound with similar retention time.

All of the retention times of the standard compounds are close to the suspects, from a couple of seconds (Carbamazepine-10,11-epoxide and Atrazine-desisopropyl), to 12 s for Efavirenz. The exception is Chlorpyrifos, which is eluting more than one minute after its closest neighbour Simvastatin. As previously discussed, more late eluting compounds in the mix would have been desirable. If this had been implemented, there would probably have been a standard compound eluting closer in time to the suspect. However, it does not appear as the closer RT between suspect and standard give lower errors, but rather the opposite. This approach is overall the worst of the ones tested in this study, as can be seen concluded by

looking at the high errors. Most probably, the response factor which is used for the semi-quantification depends on many factors that are not accounted for by this approach.

Table 3. *Structural, hydrophobicity and basicity comparison of the suspect compound and its most similar standard. The mean error of the suspect is also compared with the maximum expected error. logP and pK_a were calculated in ChemAxon, similarity score and maximum expected error were from http://dsfp.chem.uoa.gr/semiquantification/.*

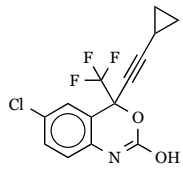
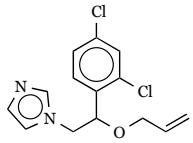
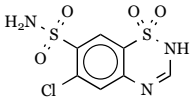
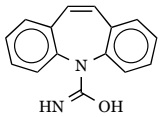
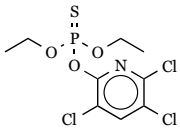
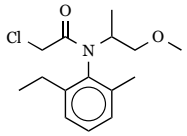
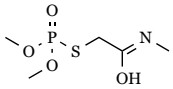
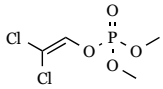
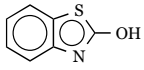
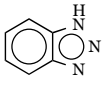
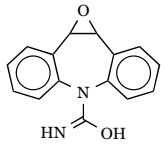
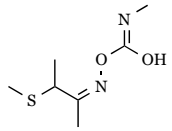
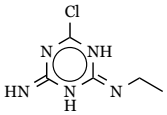
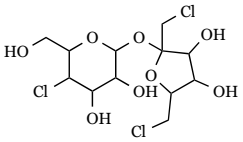
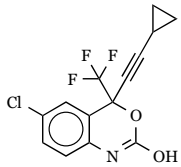
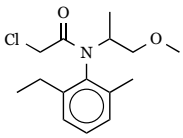
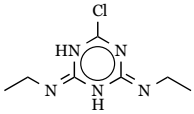
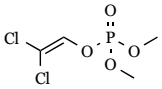
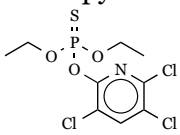
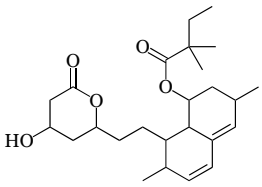
Suspect compound	Most similar standard compound
Efavirenz  Mean error: 891.2 log*P*: 4.5 pK_a: -1.5	Imazalil  Similarity score: 9.5 % Maximum expected error: 18.6 log*P*: 3.8 pK_a: 6.5
Chlorothiazide  Mean error: 562.0 log*P*: -0.44 pK_a: 1.2	Carbamazepine  Similarity score: 12.0 % Maximum expected error: 15.1 log*P*: 2.8 pK_a: -3.8
Chlorpyrifos  Mean error: 666.3 log*P*: 4.8 pK_a: -4.2	Metolachlor  Similarity score: 12.5 % Maximum expected error: 15.1 log*P*: 3.5 pK_a: -4.1
Omethoate  Mean error: 257.5 log*P*: -0.55 pK_a: NA	Dichlorvos  Similarity score: 21.0 % Maximum expected error: 11.9 log*P*: 1.4 pK_a: NA
2-Hydroxybenzothiazole  Mean error: 44.4 log*P*: 2.5 pK_a: -1.3	Benzotriazole  Similarity score: 35.0 % Maximum expected error: 9.2 log*P*: 1.3 pK_a: 0.22

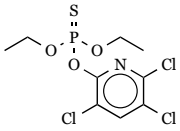
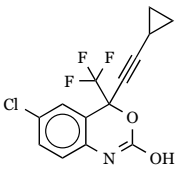
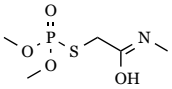
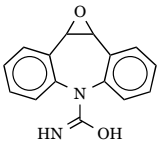
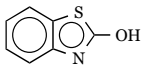
Table 4. *Comparison of structure as well as RT for the suspect compound and the closest eluting standard. *Expert pK_a estimation for Carbamazepine-10,11-epoxide, Dichlorvos and Simvastatin was 1 to 4 for Carbamazepine-10,11-epoxide, around -7 for Simvastatin and approximately 10 units lower for Dichlorvos.*

Suspect compound	Closest eluting standard
Carbamazepine-10,11-epoxide  Mean error: 1994.5 RT: 7.74 min log*P*: 2.0 pK_a: NA* Mean RF: 2.3e+17	Butocarboxim  RT: 7.72 min log*P*: 1.3 pK_a: 1.33 Mean RF: 1.1e+14
Atrazine-desisopropyl  Mean error: 635.0 RT: 5.19 min log*P*: 0.39 pK_a: 4.4 Mean RF: 2.7e+16	Sucralose $[M+Na]^+$  RT: 5.17 min log*P*: -0.47 pK_a: – Mean RF: 4.4e+13
Efavirenz  Mean error: 231.5 RT: 13.57 min log*P*: 4.5 pK_a: -1.5 Mean RF: 2.0e+14	Metolachlor  RT: 13.37 min log*P*: 3.5 pK_a: -4.1 Mean RF: 4.7e+16
Simazine  Mean error: 145.2 RT: 8.77 min log*P*: 1.8 pK_a: 4.2 Mean RF: 6.4e+16	Dichlorvos  RT: 8.91 min log*P*: 1.4 pK_a: NA* Mean RF: 4.5e+14
Chlorpyrifos  Mean error: 226.5 RT: 16.58 min log*P*: 4.8 pK_a: -4.2 Mean RF: 7.3e+13	Simvastatin  RT: 15.54 min log*P*: 4.5 pK_a: NA* Mean RF: 2.0e+16

4.3.4. Ionisation Efficiency Approach

This approach did not use the response factor of any specific standard compound to quantify the suspect, but rather predicted the ionisation efficiency. The prediction was based on 2D molecular descriptors, similar as to a fingerprint of the molecule. These descriptors aim to give similar results as 3D optimisation, but at a realistic time scale. As the model was developed, it was noticed that structural parameters were of most importance.[23] Among the 15 key features was the number of hydrogen bonds and the number of nitrogen atoms in the molecule. These can be related to hydrophobicity and basicity, respectively,[23] which are the two main properties discussed in this thesis. As seen in table 5, log*P* and pK_a of the compounds with the highest errors are displayed. Theoretically, higher log*P* values should yield lower prediction error; however, this does not seem to be the case for these substances. This might be because the relatively high log*P* values are overshadowed by the very weak basicity exhibited by the compounds.

Table 5. *The substances with the highest error for the ionisation efficiency approach, with each compounds respective RF, predicted RF, logP, pK_a and mean error. logP and pK_a was calculated in ChemAxon.*
**Expert pK_a estimation for Omethoate and Carbamazepine-10,11-epoxide, was between 1 to 4 for Carbamazepine-10,11-epoxide and approximately -2 for Omethoate.*

Suspect compound	log (Real RF)	log (Predicted RF)	log*P*	pK_a	Mean error
Chlorpyrifos 	13.9	16.0	4.8	-4.2	144.5
Efavirenz 	14.3	15.8	4.5	-1.5	31.5
Omethoate 	17.1	15.7	-0.55	NA*	24.0
Carbamazepine-10,11-epoxide 	17.4	16.8	2.0	NA*	20.8
2-Hydroxy benzothiazole 	14.4	15.5	2.5	-1.3	12.7

In figure 11, the logarithm of the predicted response factor is plotted against the logarithm of the real response factor. As seen, the data points all follow some trendline, albeit the correlation is somewhat better for the standard substances, i.e. the left-hand plot. Obviously,

the error is increasing with the difference between the predicted and real RF, as seen in table 5.

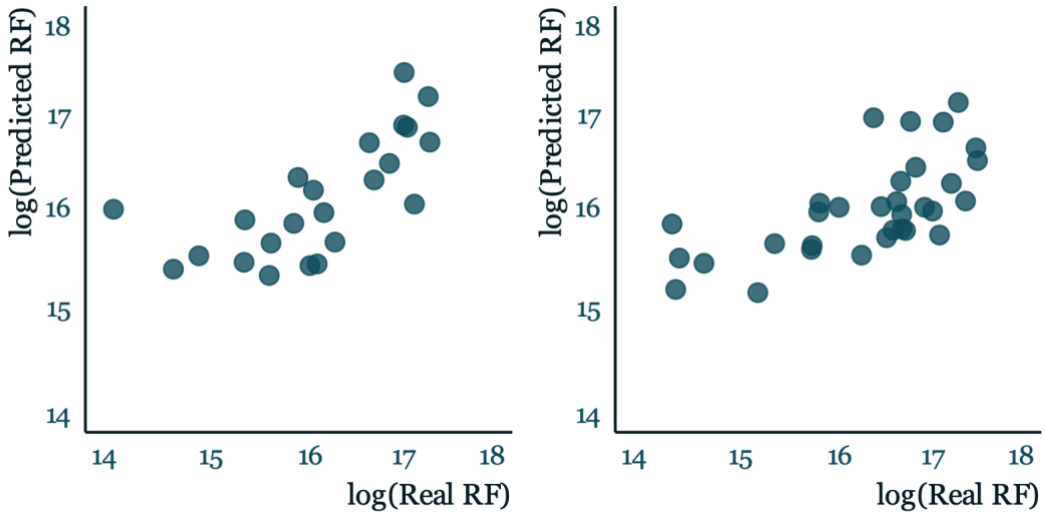


Figure 11. *Plots showing the predicted response factor vs the real response factor. The left is for the standard compounds and the right plot is for the suspect compounds.*

4.4. Stability

The results from the stability testing gave somewhat inconclusive results, as seen when comparing the two normalisation approaches in figure 6. The different ways to normalise the peak areas give quite opposite results, and neither of the assumptions that were made were necessarily correct. Actually, the measurements from week 1 would suggest that the instrument in fact was not stable, and thus, the other normalisation approach would appear to be better, i.e. the right-hand plot. However, the stability of Metformin in water is unknown, and thus the normalisation against Metformin might not be valid. Therefore, the stability needs additional investigation and quantification based on analytical standards.

5. Conclusion

Development of new analytical techniques is a prerequisite for increasing environmental awareness. For this purpose, non-targeted screening of environmental samples is being employed more and more. This is especially true in the search for emerging contaminants in water samples. Though semi-quantification methods for non-targeted analysis are available, these are not yet standardised. A desire to validate semi-quantitative non-targeted analysis led the NORMAN Network to arrange an interlaboratory comparison on the subject, organised by Stockholm University and University of Athens. As part of this thesis, four semi-quantification approaches proposed for collaborative trial were tested.

The results show quite clearly that the closest eluting standard approach is the worst performing strategy, no matter which criteria it is based on. This approach gave both the highest mean and median error, with up to 2000-fold maximum error. It also gave unacceptable high errors (over 10-fold) for the largest number of compounds. This is quite unfortunate, since this strategy was the only one of the four that could be easily used in true non-targeted screening as it does not require a tentative structure.

Comparing the three other strategies is not as straight forward. The similar compound and parent compound approaches present the lowest median error, with the parent approach having slightly lower median error. The parent compound strategy was also able to give satisfying predictions for all compounds except three. However, this approach does not work for all compound, as it is applicable to degradation products. The similar compound method is more universal, as most substances are to some extent similar to another substance. Still, the maximum errors for this approach is up to 1000-fold, which thereby gives quite high mean errors. Across the samples, this approach had undesirable results for 7-8 compounds. The ionisation efficiency method on the other hand exhibited the lowest mean error of all approaches, and the second lowest number of compounds with over 10-fold error. However, the median error was the second worst for this approach. Though the *IE* approach seems to be the most promising method, the fact that it only works for protonated or deprotonated molecules is unfortunate, as it means that it will not work for all compounds.

In conclusion, no one approach appears to be sufficient as a universal semi-quantification approach, but rather different strategies can to be used in parallel where applicable.

6. Future Challenges

The results presented in this thesis need to, and will be, further investigated. In the upcoming NORMAN collaborative trial, the results from different laboratories will be compared, and two additional semi-quantification strategies will be tested. More data input will, hopefully, show clearer trends, and thus facilitate evaluation of the approaches.

For future work, further development of ionisation efficiency prediction models is encouraged. Especially, further investigation in predicting the low sensitivity substances is needs to improve prediction accuracy. Additionally, adding the prediction possibility for sodium or ammonium adducts is beneficial to widen the application range method.

7. Acknowledgements

I would like to say special thanks to my supervisor Anneli Kruve, first of all for giving me the amazing opportunity to be a part of the NORMAN collaborative trial. Secondly, for always being willing to help me and explain when I was in doubt. Thirdly, for all the ice cream along this project.

I would also like to thank Claudia Möckel and Merle Plassman for all technical support and help with the instruments. Further, I thank Miklós Mohai, as his previous work helped with the compound selection, Josefine Carlsson and Ulrika Nilsson from the textile lab group for letting me raid their chemical cupboard, and Dr. Karl Kaupmees for his expert estimations of pK_a values.

Special thanks also to Emma Palm, Riccardo Costalunga, Amina Souihi and Anselm Okolo. Finally, I thank Jonathan Benskin for taking the time to be my opponent.

8. References

(1) NORMAN Network. Suspect list https://www.normandata.eu/normansusdat/.

(2) Pieke, E. N.; Granby, K.; Trier, X.; Smedsgaard, J. A Framework to Estimate Concentrations of Potentially Unknown Substances by Semi-Quantification in Liquid Chromatography Electrospray Ionization Mass Spectrometry. *Analytica Chimica Acta* **2017**, *975*, 30–41. https://doi.org/10.1016/j.aca.2017.03.054.

(3) Kruve, A. Strategies for Drawing Quantitative Conclusions from Nontargeted Liquid Chromatography–High-Resolution Mass Spectrometry Analysis. *Anal. Chem.* **2020**, *92* (7), 4691–4699. https://doi.org/10.1021/acs.analchem.9b03481.

(4) World Health Organisation. Drinking Water https://www.who.int/en/news-room/fact-sheets/detail/drinking-water.

(5) Bukola, D.; Zaid, A. Consequences of Anthropogenic Activities on Fish and the Aquatic Environment. *Poult Fish Wildl Sci* **2015**, *03* (02). https://doi.org/10.4172/2375-446X.1000138.

(6) Schwarzenbach, R. P. The Challenge of Micropollutants in Aquatic Systems. *Science* **2006**, *313* (5790), 1072–1077. https://doi.org/10.1126/science.1127291.

(7) Richardson, S. D.; Kimura, S. Y. Water Analysis: Emerging Contaminants and Current Issues. *Anal. Chem.* **2020**, *92* (1), 473–505. https://doi.org/10.1021/acs.analchem.9b05269.

(8) Kruve, A. Semi-quantitative Non-target Analysis of Water with Liquid Chromatography/High-resolution Mass Spectrometry: How Far Are We? *Rapid Commun Mass Spectrom* **2019**, *33* (S3), 54–63. https://doi.org/10.1002/rcm.8208.

(9) Bletsou, A. A.; Jeon, J.; Hollender, J.; Archontaki, E.; Thomaidis, N. S. Targeted and Non-Targeted Liquid Chromatography-Mass Spectrometric Workflows for Identification of Transformation Products of Emerging Pollutants in the Aquatic Environment. *TrAC Trends in Analytical Chemistry* **2015**, *66*, 32–44. https://doi.org/10.1016/j.trac.2014.11.009.

(10) Kiefer, K.; Müller, A.; Singer, H.; Hollender, J. New Relevant Pesticide Transformation Products in Groundwater Detected Using Target and Suspect Screening for Agricultural and Urban Micropollutants with LC-HRMS. *Water Research* **2019**, *165*, 114972. https://doi.org/10.1016/j.watres.2019.114972.

(11) European Commision. Directive of the European Parliament and of the Council: Amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Substances in the Field of Water Policy. January 31, 2012.

(12) United States Environmental Protection Agency. Toxic and priority pollutants under the clean water act https://www.epa.gov/eg/toxic-and-priority-pollutants-under-clean-water-act#priority.

(13) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505–11512. https://doi.org/10.1021/acs.est.7b02184.

(14) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.

(15) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of Novel Micropollutants in Wastewater by a Combination of Suspect and Nontarget Screening. *Environmental Pollution* **2014**, *184*, 25–32. https://doi.org/10.1016/j.envpol.2013.07.048.

(16) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse, S.; Thomaidis, N. S.; Bletsou, A.; Zwiener, C.; Ibáñez, M.; Portolés, T.; de Boer, R.; Reid, M. J.; Onghena, M.; Kunkel, U.; Schulz, W.; Guillon, A.; Noyon, N.; Leroy, G.; Bados, P.; Bogialli, S.; Stipaničev, D.; Rostkowski, P.; Hollender, J. Non-Target Screening with High-Resolution Mass Spectrometry: Critical

Review Using a Collaborative Trial on Water Analysis. *Anal Bioanal Chem* **2015**, *407* (21), 6237–6255. https://doi.org/10.1007/s00216-015-8681-7.
(17) Waters. Oasis Solid-Phase Extraction Products. Waters April 2020.
(18) NORMAN Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances https://www.norman-network.com/.
(19) NORMAN Network. NORMAN Joint Programme of Activities (JPA 2020). February 2020.
(20) Guthrie, W. F. Interlaboratory Comparisons. *Wiley StatRef: Statistics Reference Online* 7. https://doi.org/10.1002/9781118445112.stat04148.
(21) Vangel, M. G. Interlaboratory Studies. *Wiley StatRef: Statistics Reference Online* **2014**, 6. https://doi.org/10.1002/9781118445112.stat07619.
(22) Kruve, A.; Aalizadeh, R.; Alygizakis, N.; Thomaidis, N. S.; Malm, L. Interlaboratory Comparison on Strategies for Semi-Quantitative Non-Targeted LC-ESI-HRMS https://www.norman-network.net/sites/default/files/files/QA-QC%20Issues/Invitation%20letter%20JPA%202020%20semi-quant%20inter%20lab%20%28002%29.pdf.
(23) Liigand, J.; Wang, T.; Kellogg, J.; Smedsgaard, J.; Cech, N.; Kruve, A. Quantification for Non-Targeted LC/MS Screening without Standard Substances. *Sci Rep* **2020**, *10* (1), 5808. https://doi.org/10.1038/s41598-020-62573-z.
(24) Similar compound finder http://dsfp.chem.uoa.gr/semiquantification/.
(25) Dahal, U. P.; Jones, J. P.; Davis, J. A.; Rock, D. A. Small Molecule Quantification by Liquid Chromatography-Mass Spectrometry for Metabolites of Drugs and Drug Candidates. *Drug Metab Dispos* **2011**, *39* (12), 2355–2360. https://doi.org/10.1124/dmd.111.040865.
(26) Quantem Analytics https://app.quantem.co/.
(27) Rousis, N. I.; Gracia-Lor, E.; Zuccato, E.; Bade, R.; Baz-Lomba, J. A.; Castrignanò, E.; Causanilles, A.; Covaci, A.; de Voogt, P.; Hernàndez, F.; Kasprzyk-Hordern, B.; Kinyua, J.; McCall, A.-K.; Plósz, B. Gy.; Ramin, P.; Ryu, Y.; Thomas, K. V.; van Nuijs, A.; Yang, Z.; Castiglioni, S. Wastewater-Based Epidemiology to Assess Pan-European Pesticide Exposure. *Water Research* **2017**, *121*, 270–279. https://doi.org/10.1016/j.watres.2017.05.044.
(28) Kruve, A.; Kaupmees, K. Predicting ESI/MS Signal Change for Anions in Different Solvents. *Anal. Chem.* **2017**, *89* (9), 5079–5086. https://doi.org/10.1021/acs.analchem.7b00595.
(29) Kiefer, K.; Signer, H.; Hollender, J.; Kruve, A. Benchmarking of the Semi-Quantification Strategies for the Non-Targeted Screening of Micropollutants and Transformation Products in Groundwater. Unpublished.

9. Supplementary Information

Table S1. *All the compounds with their corresponding solvent, masses and concentrations. Note that for nigericin, 10,11-dihydro-10-hydroxycarbamazepine, carbamazepine-10,11-epoxide and caffeine-$^{13}C_3$ no mass for the solvent is available as they were purchased as solutions with concentration of 1mg/mL. Atrazine-desethyl-desisopropyl-2-hydroxy were prepared in a 100 mL volumetric flask and was not weighed.*

Compound	$m_{compound}$ [mg]	Solvent	$m_{solvent}$ [mg]	c [M]
Guanylurea	2.4	H_2O	10007.5	0.00158
Amitrole	8.8	MeCN	7789.6	0.01056
Histamine	11.1	MeCN	7786.2	0.01008
Chlormequat	10.7	MeCN	7786.2	0.00683
Methamidophos	17.5	MeCN	7763.4	0.01255
Vancomycin	1.7	H_2O	9975.4	0.00012
Trichlorfon	8	MeCN	7751	0.00315
Butocarboxim	10.4	MeCN	7761.6	0.00554
Dichlorvos	14.7	MeCN	7718.2	0.00677
Tylosin	10.9	MeCN	7785.7	0.00108
Rifaximin	9.7	MeCN	7783.6	0.00125
Spinosyn A	10.2	MeCN	7810.8	0.00140
Emamectin B1a	12.1	MeCN	7778.4	0.00121
Nigericin	5	DMSO:EtOH 1:1	-	0.00670
Ivermectin B1a	11.3	MeCN	7733.8	0.00131
TCMTB	10.4	MeCN	7791.6	0.00440
Atrazine	8.8	MeCN	7796.8	0.00411
Octocrylene	19.7	MeCN	7776	0.00551
Clarithromycin	10.5	MeCN	7763.2	0.00142
Aspartame	13.4	H_2O	9980.4	0.00455
Simvastatin	9.9	MeCN	7782.6	0.00239
Sucralose	13.7	MeCN	7766.8	0.00349
Irgarol	7	MeCN	7777	0.00279
Caffeine	13.4	MeCN	7771.5	0.00698
Carbamazepine	10.9	MeCN	7750.7	0.00371
Benzotriazole	14.6	MeCN	7707.3	0.01250
Metolachlor	12	MeCN	7759.2	0.00428
Imazalil	9.9	MeCN	7757.1	0.00338
Atrazine- desethyl	11.1	MeCN	7795.8	0.00596
Atrazine-desethyl-2-hydroxy	2.2	H_2O:0.1% FA 5:1	11957.8	0.00108
Atrazine-desisopropyl	9.2	MeCN	7766.3	0.00536
Atrazine-desethyl-desisopropyl	2.3	DMSO	11009.6	0.00158

Compound	$m_{compound}$ [mg]	Solvent	$m_{solvent}$ [mg]	c [M]
Atrazine-desethyl-desisopropyl-2-hydroxy	2.4	H_2O:0.1% FA 1:10	-	0.00019
Atrazine-desisopropyl-2-hydroxy	2.3	H_2O:0.1% FA 5:1	12028.8	0.00123
2-(Methylthio) benzothiazole	12.5	MeCN	7764.2	0.00848
Progesterone	11.8	MeCN	7805.5	0.00378
Butylamine	9.1	MeCN	7782.7	0.01256
Haloperidol	8	MeCN	7780.2	0.00215
Reserpine	11.6	MeCN	7771.6	0.00193
Phenazine	10.7	MeCN	7760.3	0.00601
Clotrimazole	6.8	MeCN	7776.3	0.00199
Simazine	2.7	MeOH	7733.1	0.00137
Efavirenz	10.3	MeCN	7757.7	0.00331
Adenosine	3.4	H_2O	10001.6	0.00127
Climbazole	11.9	MeCN	7766.6	0.00411
Melamine	10.2	H_2O	11982.8	0.00673
Metazachlor	9.2	MeCN	7758.6	0.00336
Chlorothiazide	14.1	MeCN	7746.4	0.00484
Metformin	10.7	H_2O	9962.3	0.00647
2-Methylbenzothiazole	12.1	MeCN	7746.6	0.00823
Benzothiazole	13.5	MeCN	7726.9	0.01016
Chlorpyrifos	10.6	MeCN	7822.1	0.00304
5-Methyl-1H-benzotriazole	9.8	MeCN	7771.5	0.00744
10,11-dihydro-10-hydroxycarbamazepine	1	MeCN	-	0.00393
Sudan I	2.3	MeCN	9315	0.00078
Ketoconazole	7.3	MeCN	7807	0.00138
5-Chlorobenzotriazole	14.2	MeCN	7763.7	0.00936
Benzotriazole-5-carboxylic acid	9.1	MeOH:HCl 10:3	10682.3	0.00452
Carbamazepine-10,11-epoxide	1	MeOH	-	0.00396
Metolachlor-ESA	2.2	MeCN	7793.4	0.00063
Metolachlor-OXA	1.7	MeCN	7814.4	0.00061
Omethoate	12.4	MeCN	7772.9	0.00588
Atrazine-2-hydroxy	9.9	MeOH:HCl 2:3	9158.3	0.00517
2-Hydroxybenzothiazole	14.4	MeCN	7781	0.00962
2-Aminobenzothiazole	11	MeCN	7783.2	0.00740

Compound	$m_{compound}$ [mg]	Solvent	$m_{solvent}$ [mg]	c [M]
Atrazine-d_5	3.1	MeCN	2323.4	0.00475
Imazalil-d_5	5.5	MeCN	3861.6	0.00370
Caffeine-$^{13}C_3$	1	MeOH	-	0.00507

Table S2. *The concentrations for the real samples, where SB1 & SB2 corresponds Laduviken samples and SB3 & SB4 corresponds to tap water samples.*

Compound	c SB1 & SB3 [M]	c SB2 & SB4 [M]
Amitrole	1.03e-07	1.03e-07
Aspartame	4.62e-08	4.62e-08
Atrazine	4.07e-08	4.07e-08
Benzotriazole	1.24e-07	1.24e-07
Butocarboxim	5.44e-08	5.44e-08
Caffeine	6.84e-08	6.84e-08
Carbamazepine	3.65e-08	3.65e-08
Chlormequat	6.72e-08	6.72e-08
Clarithromycin	1.40e-08	1.40e-08
Dichlorvos	6.66e-08	6.66e-08
Emamectin B1a	1.20e-08	1.20e-08
Guanylurea	7.85e-08	7.85e-08
Histamine	9.84e-08	9.84e-08
Imazalil	3.31e-08	3.31e-08
Irgarol	2.74e-08	2.74e-08
Ivermectin B1a	1.29e-08	1.29e-08
Methamidophos	1.24e-07	1.24e-07
Metolachlor	4.20e-08	4.20e-08
Nigericin	1.32e-08	1.32e-08
Octocrylene	2.59e-08	2.59e-08
Rifaximin	1.22e-08	1.22e-08
Simvastatin	2.32e-08	2.32e-08
Spinosyn A	1.39e-08	1.39e-08
Sucralose	3.43e-08	3.43e-08
TCMTB	4.35e-08	4.34e-08
Trichlorfon	3.12e-08	3.12e-08
Tylosin	1.08e-08	1.08e-08
Vancomycin	5.91e-09	5.91e-09
Atrazine- desethyl	2.23e-07	1.11e-07
Atrazine-desethyl-2-hydroxy	3.18e-07	1.59e-07
Atrazine-desisopropyl	2.07e-07	1.04e-07
Atrazine-desethyl-desisopropyl	3.98e-07	1.99e-07
Atrazine-desethyl-desisopropyl-2-hydroxy	4.79e-07	2.40e-07
Atrazine-desisopropyl-2-hydroxy	3.62e-07	1.81e-07
2-(Methylthio)benzothiazole	2.95e-07	1.47e-07
Progesterone	1.36e-07	6.81e-08
Butylamine	1.02e-06	5.09e-07
Haloperidol	8.58e-08	4.29e-08
Reserpine	7.19e-08	3.60e-08

Compound	*c* SB1 & SB3 [M]	*c* SB2 & SB4 [M]
Phenazine	7.74e-08	3.87e-08
Clotrimazole	1.72e-07	8.59e-08
Simazine	2.77e-07	1.38e-07
Efavirenz	1.57e-06	7.85e-07
Adenosine	3.05e-07	1.52e-07
Climbazole	5.29e-08	2.65e-08
Melamine	9.56e-07	4.78e-07
Metazachlor	1.17e-07	5.83e-08
Chlorothiazide	2.30e-06	1.15e-06
Metformin	5.97e-07	2.98e-07
2-Methylbenzothiazole	7.31e-07	3.65e-07
Benzothiazole	4.90e-06	2.45e-06
Chlorpyrifos	2.94e-06	1.47e-06
5-Methyl-1H-benzotriazole	2.78e-07	1.39e-07
10,11-dihydro-10-hydroxycarbamazepine	1.32e-07	6.58e-08
Sudan I	1.04-07	5.18e-08
Ketoconazole	6.50e-07	3.25e-07
5-Chlorobenzotriazole	7.83e-07	3.92e-07
Benzotriazole-5-carboxylic acid	2.17e-06	1.08e-06
Carbamazepine-10,11-epoxide	7.09e-08	3.54e-08
Metolachlor-ESA	1.56e-06	7.79e-07
Metolachlor-OXA	1.49e-06	7.47e-07
Omethoate	2.57e-07	1.29e-07
Atrazine-2-hydroxy	9.98e-08	4.99e-08
2-Hydroxybenzothiazole	3.96e-07	1.98e-07
2-Aminobenzothiazole	2.66e-07	1.33e-07
Atrazine-d_5	4.15e-08	4.15e-08
Caffeine-$^{13}C_3$	3.36e-08	3.36e-08
Imazalil-d_5	3.09e-08	3.09e-08

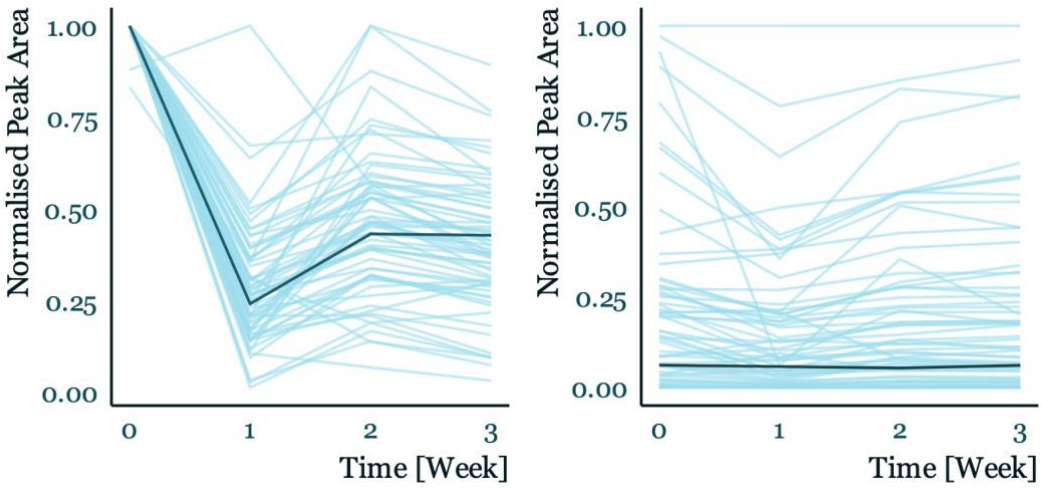


Figure S1. *The stability plots with data from week 1 included. As seen, something happened with the instrumental stability in week 1. This is clearest in the left plot.*

Table S3. *Comparison of the predicted concentrations for all approaches and the real concentration for all suspect compounds in all four samples.*

Compound	Sample	c_{real}	Closest Eluting Std c_{pred}	Ionisation Efficiency c_{pred}	Parent Compound c_{pred}	Similar Compound c_{pred}
10,11-Dihydro-10-hydroxycarbamazepine	SB1	1.32e-07	3.22e-08	7.04e-07	NA	NA
2-(Methylthio)benzothiazole	SB1	2.95e-07	8.25e-07	1.98e-07	8.25e-07	8.25e-07
2-Aminobenzothiazole	SB1	2.67e-07	1.97e-06	2.28e-06	6.13e-06	6.13e-06
2-Hydroxybenzothiazole	SB1	3.96e-07	8.54e-07	3.39e-08	4.79e-08	1.05e-08
2-Methylbenzothaizole	SB1	7.31e-07	4.18e-08	4.07e-07	NA	NA
5-Chlorobenzotriazole	SB1	7.83e-07	9.55e-08	1.07e-06	4.05e-07	4.05e-07
5-Methyl-1H-benzotriazole	SB1	2.78e-07	5.48e-08	1.73e-06	1.03e-06	1.03e-06
Adenosine	SB1	3.05e-07	1.36e-06	1.57e-06	NA	2.78e-08
Atrazine-2-hydroxy	SB1	9.98e-08	1.45e-06	1.06e-06	1.42e-07	1.42e-07
Atrazine-desethyl	SB1	2.23e-07	4.81e-08	7.69e-07	1.29e-07	1.29e-07
Atrazine-desethyl-2-hydroxy	SB1	3.18e-07	2.60e-06	2.01e-06	1.43e-07	1.43e-07
Atrazine-desethyl-desisopropyl	SB1	3.98e-07	1.76e-07	5.06e-07	2.99e-08	9.72e-07
Atrazine-desethyl-desisopropyl-2-hydroxy	SB1	4.79e-07	5.64e-08	7.89e-08	3.10e-09	1.01e-07
Atrazine-desisopropyl	SB1	2.07e-07	1.48e-04	5.48e-07	8.14e-08	8.14e-08
Atrazine-desisopropyl-2-hydroxy	SB1	3.62e-07	4.36e-06	2.73e-06	2.39e-07	2.39e-07
Benzothiazole	SB1	4.90e-06	2.43e-05	2.49e-06	4.92e-06	1.08e-06
Benzotriazole-5-carboxylic acid	SB1	2.17e-06	1.78e-07	2.32e-06	NA	NA
Carbamazepine-10,11-epoxide	SB1	7.09e-08	1.33e-04	1.38e-06	NA	NA
Chlorothiazide	SB1	2.30e-06	3.09e-08	3.74e-07	NA	4.29e-09
Chlorpyrifos	SB1	2.94e-06	2.40e-08	2.35e-08	NA	5.48e-09

Compound	Sample	c_{real}	Closest Eluting Std c_{pred}	Ionisation Efficiency c_{pred}	Parent Compound c_{pred}	Similar Compound c_{pred}
Climbazole	SB1	5.29e-08	1.64e-07	5.25e-07	NA	8.90e-08
Clotrimazole	SB1	1.72e-07	5.13e-07	1.20e-06	NA	3.96e-07
Efavirenz	SB1	1.57e-06	7.48e-09	5.19e-08	NA	1.90e-09
Haloperidol	SB1	8.58e-08	8.79e-08	1.21e-07	NA	8.79e-08
Ketoconazole	SB1	6.50e-07	2.00e-06	1.02e-06	NA	4.52e-07
Metazachlor	SB1	1.17e-07	1.88e-07	1.03e-06	NA	3.94e-07
Metformin	SB1	5.97e-07	2.25e-05	4.74e-06	NA	1.21e-05
Omethoate	SB1	2.57e-07	4.46e-06	5.75e-06	NA	6.86e-05
Phenazine	SB1	7.74e-08	4.94e-08	1.81e-07	NA	2.63e-08
Progesterone	SB1	1.36e-07	2.84e-08	1.32e-07	NA	2.84e-08
Reserpine	SB1	7.19e-08	4.04e-08	4.86e-08	NA	8.60e-08
Simazine	SB1	2.77e-07	4.05e-05	6.68e-07	NA	2.57e-07
Sudan I	SB1	1.04e-07	2.19e-07	2.57e-08	NA	1.81e-08
10,11-Dihydro-10-hydroxycarbamazepine	SB2	6.58e-08	1.65e-08	3.38e-07	NA	NA
2-(Methylthio)benzothiazole	SB2	1.47e-07	3.90e-07	9.54e-08	3.90e-07	3.90e-07
2-Aminobenzothiazole	SB2	1.33e-07	9.05e-07	1.05e-06	2.77e-06	2.77e-06
2-Hydroxybenzothiazole	SB2	1.98e-07	4.04e-07	1.30e-08	1.82e-08	4.24e-09
2-Methylbenzothaizole	SB2	3.65e-07	2.11e-08	1.92e-07	NA	NA
5-Chlorobenzotriazole	SB2	3.92e-07	5.62e-08	5.69e-07	2.25e-07	2.25e-07
5-Methyl-1H-benzotriazole	SB2	1.39e-07	2.81e-08	8.32e-07	5.15e-07	5.15e-07
Adenosine	SB2	1.52e-07	7.85e-07	8.39e-07	NA	1.59e-08
Atrazine-2-hydroxy	SB2	4.99e-08	7.35e-07	5.40e-07	7.64e-08	7.64e-08
Atrazine-desethyl	SB2	1.11e-07	1.90e-08	NA	5.07e-08	5.07e-08
Atrazine-desethyl-2-hydroxy	SB2	1.59e-07	1.39e-06	9.94e-07	7.53e-08	7.53e-08

Compound	Sample	c_{real}	Closest Eluting Std c_{pred}	Ionisation Efficiency c_{pred}	Parent Compound c_{pred}	Similar Compound c_{pred}
Atrazine-desethyl-desisopropyl	SB2	1.99e-07	9.14e-08	2.36e-07	1.49e-08	4.35e-07
Atrazine-desethyl-desisopropyl-2-hydroxy	SB2	2.40e-07	2.74e-08	3.53e-08	1.48e-09	4.34e-08
Atrazine-desisopropyl	SB2	1.04e-07	3.45e-05	2.60e-07	4.09e-08	4.09e-08
Atrazine-desisopropyl-2-hydroxy	SB2	1.81e-07	2.38e-06	1.39e-06	1.29e-07	1.29e-07
Benzothiazole	SB2	2.45e-06	1.07e-05	1.16e-06	2.27e-06	5.28e-07
Benzotriazole-5-carboxylic acid	SB2	1.08e-06	9.31e-08	1.12e-06	NA	NA
Carbamazepine-10,11-epoxide	SB2	3.54e-08	8.09e-05	6.89e-07	NA	NA
Chlorothiazide	SB2	1.15e-06	1.47e-08	1.64e-07	NA	1.92e-09
Chlorpyrifos	SB2	1.47e-06	3.29e-09	1.02e-08	NA	2.48e-09
Climbazole	SB2	2.65e-08	8.11e-08	2.56e-07	NA	4.55e-08
Clotrimazole	SB2	8.59e-08	2.53e-07	5.84e-07	NA	1.90e-07
Efavirenz	SB2	7.85e-07	3.49e-09	2.32e-08	NA	9.05e-10
Haloperidol	SB2	4.29e-08	4.52e-08	6.00e-08	NA	4.52e-08
Ketoconazole	SB2	3.25e-07	1.01e-06	4.80e-07	NA	2.20e-07
Metazachlor	SB2	5.83e-08	9.55e-08	4.63e-07	NA	1.83e-07
Metformin	SB2	2.98e-07	1.27e-05	2.81e-06	NA	7.52e-06
Omethoate	SB2	1.29e-07	2.22e-06	2.85e-06	NA	3.22e-05
Phenazine	SB2	3.87e-08	2.56e-08	8.89e-08	NA	1.28e-08
Progesterone	SB2	6.81e-08	1.46e-08	6.54e-08	NA	1.46e-08
Reserpine	SB2	3.60e-08	1.78e-08	2.14e-08	NA	3.86e-08
Simazine	SB2	1.38e-07	1.85e-05	3.28e-07	NA	1.32e-07
Sudan I	SB2	5.18e-08	3.05e-08	1.16e-08	NA	8.01e-09

Compound	Sample	c_{real}	Closest Eluting Std c_{pred}	Ionisation Efficiency c_{pred}	Parent Compound c_{pred}	Similar Compound c_{pred}
10,11-Dihydro-10-hydroxycarbamazepine	SB3	1.32e-07	3.40e-08	7.01e-07	NA	NA
2-(Methylthio)benzothiazole	SB3	2.95e-07	8.03e-07	1.81e-07	8.74e-07	8.74e-07
2-Aminobenzothiazole	SB3	2.67e-07	1.97e-06	2.18e-06	6.67e-06	6.67e-06
2-Hydroxybenzothiazole	SB3	3.96e-07	7.32e-07	2.84e-08	4.47e-08	8.43e-09
2-Methylbenzothaizole	SB3	7.31e-07	4.23e-08	3.85e-07	NA	NA
5-Chlorobenzotriazole	SB3	7.83e-07	1.02e-07	1.11e-06	4.05e-07	4.05e-07
5-Methyl-1H-benzotriazole	SB3	2.78e-07	5.42e-08	1.64e-06	9.51e-07	9.51e-07
Adenosine	SB3	3.05e-07	1.35e-06	1.53e-06	NA	2.59e-08
Atrazine-2-hydroxy	SB3	9.98e-08	1.49e-06	1.02e-06	1.25e-07	1.25e-07
Atrazine-desethyl	SB3	2.23e-07	3.99e-08	NA	9.20e-08	9.20e-08
Atrazine-desethyl-2-hydroxy	SB3	3.18e-07	2.36e-06	1.76e-06	1.12e-07	1.12e-07
Atrazine-desethyl-desisopropyl	SB3	3.98e-07	1.73e-07	4.83e-07	2.54e-08	9.48e-07
Atrazine-desethyl-desisopropyl-2-hydroxy	SB3	4.79e-07	4.81e-08	6.62e-08	2.28e-09	8.51e-08
Atrazine-desisopropyl	SB3	2.07e-07	1.54e-04	5.33e-07	7.27e-08	7.27e-08
Atrazine-desisopropyl-2-hydroxy	SB3	3.62e-07	4.33e-06	2.59e-06	2.05e-07	2.05e-07
Benzothiazole	SB3	4.90e-06	2.61e-05	2.59e-06	5.76e-06	1.09e-06
Benzotriazole-5-carboxylic acid	SB3	2.17e-06	1.83e-07	2.39e-06	NA	NA
Carbamazepine-10,11-epoxide	SB3	7.09e-08	1.36e-04	1.32e-06	NA	NA
Chlorothiazide	SB3	2.30e-06	3.52e-08	4.27e-07	NA	4.56e-09
Chlorpyrifos	SB3	2.94e-06	2.11e-08	1.92e-08	NA	4.37e-09
Climbazole	SB3	5.29e-08	1.75e-07	5.04e-07	NA	8.83e-08

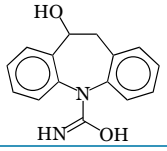
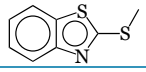
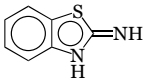
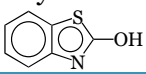
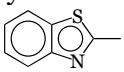
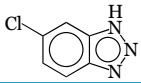
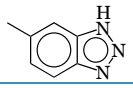
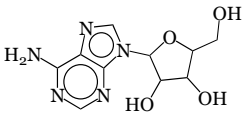
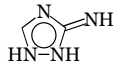
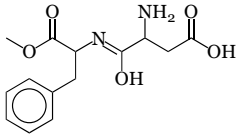
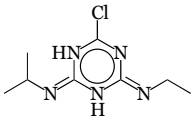
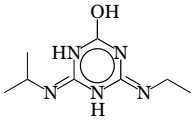
Compound	Sample	c_{real}	Closest Eluting Std c_{pred}	Ionisation Efficiency c_{pred}	Parent Compound c_{pred}	Similar Compound c_{pred}
Clotrimazole	SB3	1.72e-07	5.65e-07	1.17e-06	NA	4.07e-07
Efavirenz	SB3	1.57e-06	7.62e-09	5.04e-08	NA	1.81e-09
Haloperidol	SB3	8.58e-08	9.59e-08	1.13e-07	NA	8.95e-08
Ketoconazole	SB3	6.50e-07	2.03e-06	9.08e-07	NA	4.29e-07
Metazachlor	SB3	1.17e-07	2.01e-07	1.02e-06	NA	3.83e-07
Metformin	SB3	5.97e-07	2.09e-05	4.12e-06	NA	1.12e-05
Omethoate	SB3	2.57e-07	4.36e-06	5.39e-06	NA	6.42e-05
Phenazine	SB3	7.74e-08	5.00e-08	1.68e-07	NA	2.49e-08
Progesterone	SB3	1.36e-07	3.06e-08	1.34e-07	NA	2.78e-08
Reserpine	SB3	7.19e-08	4.17e-08	4.37e-08	NA	8.07e-08
Simazine	SB3	2.77e-07	4.17e-05	6.23e-07	NA	2.28e-07
Sudan I	SB3	1.04e-07	2.11e-07	2.15e-08	NA	1.63e-08
10,11-Dihydro-10-hydroxycarbamazepine	SB4	6.58e-08	1.80e-08	5.17e-07	NA	NA
2-(Methylthio)benzothiazole	SB4	1.47e-07	3.85e-07	1.21e-07	4.18e-07	4.18e-07
2-Aminobenzothiazole	SB4	1.33e-07	9.79e-07	1.52e-06	3.31e-06	3.31e-06
2-Hydroxybenzothiazole	SB4	1.98e-07	3.58e-07	1.97e-08	2.18e-08	4.29e-09
2-Methylbenzothaizole	SB4	3.65e-07	2.16e-08	2.74e-07	NA	NA
5-Chlorobenzotriazole	SB4	3.92e-07	4.26e-08	6.54e-07	1.76e-07	1.76e-07
5-Methyl-1H-benzotriazole	SB4	1.39e-07	2.50e-08	1.06e-06	4.56e-07	4.56e-07
Adenosine	SB4	1.52e-07	7.02e-07	1.14e-06	NA	1.61e-08
Atrazine-2-hydroxy	SB4	4.99e-08	7.61e-07	7.31e-07	6.48e-08	6.48e-08
Atrazine-desethyl	SB4	1.11e-07	1.93e-08	NA	4.52e-08	4.52e-08
Atrazine-desethyl-2-hydroxy	SB4	1.59e-07	1.24e-06	1.29e-06	5.92e-08	5.92e-08
Atrazine-desethyl-desisopropyl	SB4	1.99e-07	7.90e-08	3.11e-07	1.17e-08	3.88e-07

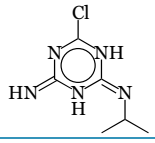
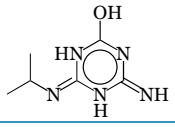
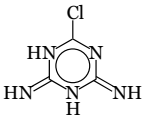
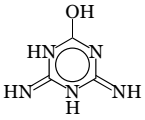
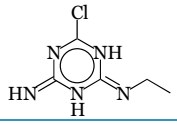
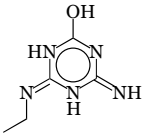
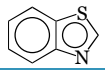
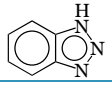
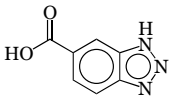
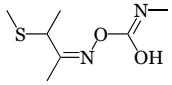
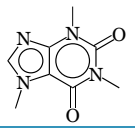
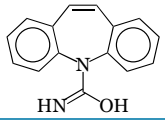
Compound	Sample		Closest Eluting Std	Ionisation Efficiency	Parent Compound	Similar Compound
		c_{real}	c_{pred}	c_{pred}	c_{pred}	c_{pred}
Atrazine-desethyl-desisopropyl-2-hydroxy	SB4	2.40e-07	2.46e-08	4.79e-08	1.18e-09	3.89e-08
Atrazine-desisopropyl	SB4	1.04e-07	7.81e-05	3.77e-07	3.74e-08	3.74e-08
Atrazine-desisopropyl-2-hydroxy	SB4	1.81e-07	2.36e-06	1.98e-06	1.13e-07	1.13e-07
Benzothiazole	SB4	2.45e-06	1.30e-05	1.82e-06	2.86e-06	5.63e-07
Benzotriazole-5-carboxylic acid	SB4	1.08e-06	8.57e-08	1.60e-06	NA	NA
Carbamazepine-10,11-epoxide	SB4	3.54e-08	6.68e-05	8.99e-07	NA	NA
Chlorothiazide	SB4	1.15e-06	1.50e-08	2.60e-07	NA	1.89e-09
Chlorpyrifos	SB4	1.47e-06	7.43e-09	9.43e-09	NA	1.70e-09
Climbazole	SB4	2.65e-08	8.16e-08	3.23e-07	NA	4.93e-08
Clotrimazole	SB4	8.59e-08	1.83e-07	5.19e-07	NA	1.28e-07
Efavirenz	SB4	7.85e-07	2.75e-09	2.55e-08	NA	7.82e-10
Haloperidol	SB4	4.29e-08	3.96e-08	6.27e-08	NA	4.42e-08
Ketoconazole	SB4	3.25e-07	6.77e-07	4.10e-07	NA	1.71e-07
Metazachlor	SB4	5.83e-08	9.64e-08	6.74e-07	NA	2.03e-07
Metformin	SB4	2.98e-07	1.33e-05	3.66e-06	NA	7.61e-06
Omethoate	SB4	1.29e-07	2.25e-06	3.91e-06	NA	3.41e-05
Phenazine	SB4	3.87e-08	2.55e-08	1.18e-07	NA	1.23e-08
Progesterone	SB4	6.81e-08	1.45e-08	8.85e-08	NA	1.46e-08
Reserpine	SB4	3.60e-08	1.42e-08	2.01e-08	NA	3.03e-08
Simazine	SB4	1.38e-07	2.07e-05	4.26e-07	NA	1.15e-07
Sudan I	SB4	5.18e-08	6.50e-08	8.97e-09	NA	4.87e-09

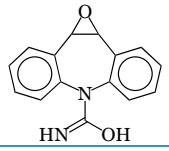
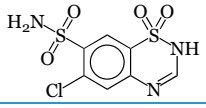
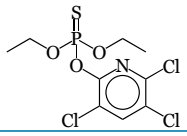
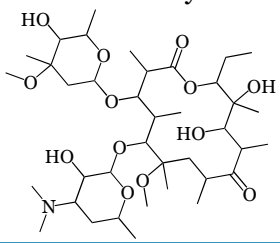
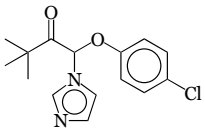
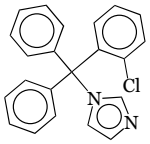
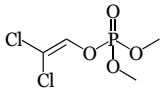
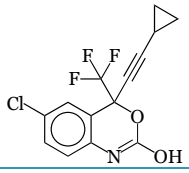
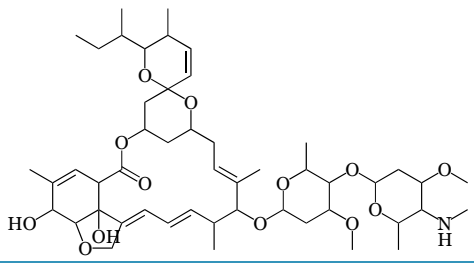
Table S4. *All suspect compound displayed together with their respective closest eluting-, parent-, and most similar compound.*

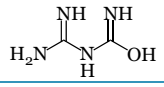
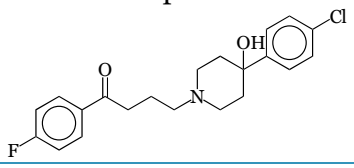
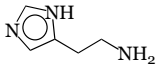
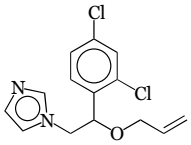
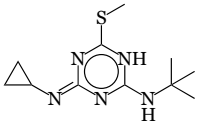
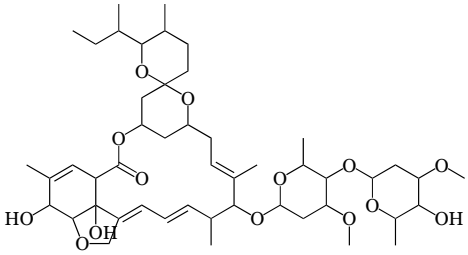
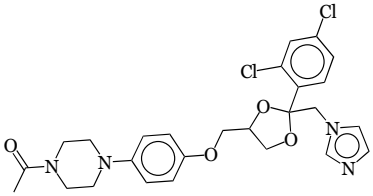
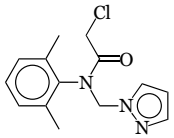
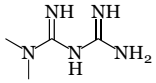
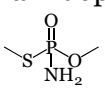
Suspect compound	Closest eluting compound	Parent compound	Most similar compound
10,11-Dihydro-10-hydroxycarbamazepine	Imazalil	NA	NA
2-(Methylthio)benzothiazole	TCMTB	TCMTB	TCMTB
2-Aminobenzothiazole	Vancomycin $[M]^{2+}$	TCMTB	TCMTB
2-Hydroxybenzothiazole	Butocarboxim	TCMTB	Benzotriazole
2-Methylbenzothaizole	Spinosyn A	NA	NA
5-Chlorobenzotriazole	Tylosin	Benzotriazole	Benzotriazole
5-Methyl-1H-benzotriazole	Imazalil	Benzotriazole	Benzotriazole
Adenosine	Amitrole	NA	Imazalil
Atrazine-2-hydroxy	Vancomycin $[M]^{2+}$	Atrazine	Atrazine
Atrazine-desethyl	Imazalil	Atrazine	Atrazine
Atrazine-desethyl-2-hydroxy	Amitrole	Atrazine	Atrazine
Atrazine-desethyl-desisopropyl	Methamidophos	Atrazine	Guanlylurea
Atrazine-desethyl-desisopropyl-2-hydroxy	Amitrole	Atrazine	Guanlylurea
Atrazine-desisopropyl	Sucralose $[M+Na]^{+}$	Atrazine	Atrazine
Atrazine-desisopropyl-2-hydroxy	Amitrole	Atrazine	Atrazine
Benzothiazole	Dichlorvos	TCMTB	Benzotriazole
Benzotriazole-5-carboxylic acid	Aspartame	NA	NA
Carbamazepine-10,11-epoxide	Butocarboxim	NA	NA
Chlorothiazide	Aspartame	NA	Carbamazepine
Chlorpyrifos	Simvastatin	NA	Metolachlor
Climbazole	Clarithromycin	NA	Imazalil
Clotrimazole	Clarithromycin	NA	Carbamazepine
Efavirenz	Metolachlor	NA	Imazalil
Haloperidol	Imazalil	NA	Imazalil
Ketoconazole	Tylosin	NA	Imazalil
Metazachlor	Emamectin B1a	NA	Metolachlor
Metformin	Guanlylurea	NA	Caffeine
Omethoate	Vancomycin $[M]^{2+}$	NA	Dichlorvos
Phenazine	Atrazine	NA	Carbamazepine
Progesterone	Metolachlor	NA	Metolachlor
Reserpine	Clarithromycin	NA	Metolachlor
Simazine	Dichlorvos	NA	Atrazine
Sudan I	Simvastatin	NA	Carbamazepine

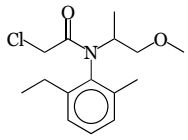
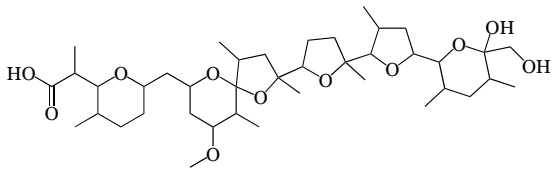
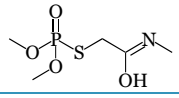
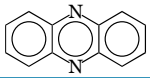
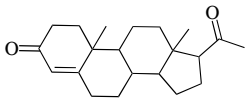
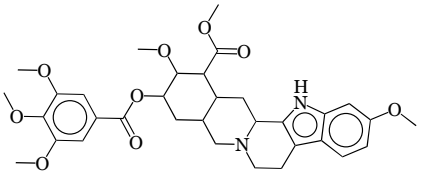
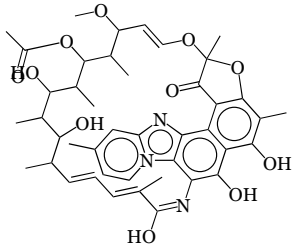
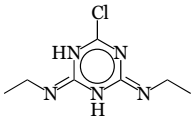
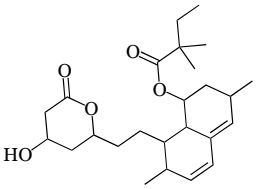
Table S5. *The structures of all the compounds that were visible in the analysis.*

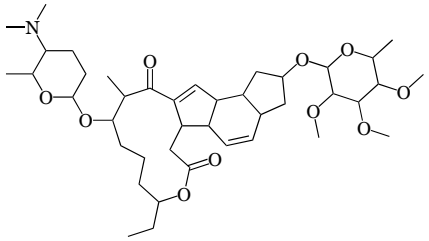
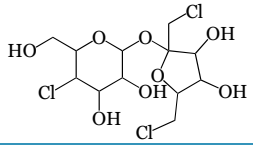
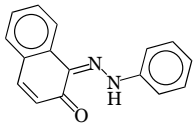
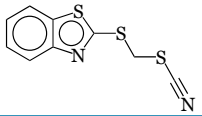
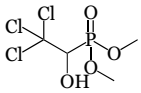
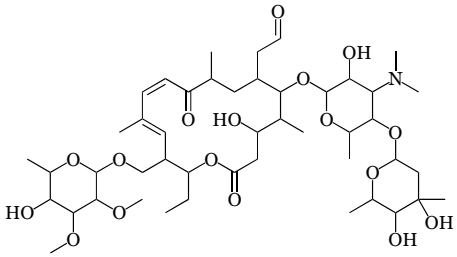
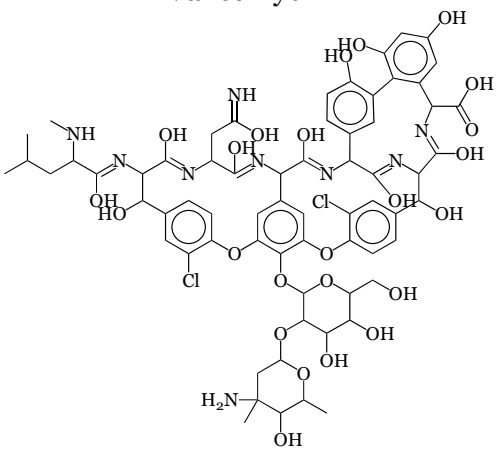
Structure
10,11-Dihydro-10-hydroxycarbamazepine 
2-(Methylthio)benzothiazole 
2-Aminobenzothiazole 
2-Hydroxybenzothiazole 
2-Methylbenzothaizole 
5-Chlorobenzotriazole 
5-Methyl-1H-benzotriazole 
Adenosine 
Amitrole 
Aspartame 
Atrazine 
Atrazine-2-hydroxy 

Structure
Atrazine-desethyl 
Atrazine-desethyl-2-hydroxy 
Atrazine-desethyl-desisopropyl 
Atrazine-desethyl-desisopropyl-2-hydroxy 
Atrazine-desisopropyl 
Atrazine-desisopropyl-2-hydroxy 
Benzothiazole 
Benzotriazole 
Benzotriazole-5-carboxylic acid 
Butocarboxim 
Caffeine 
Carbamazepine 

Structure
Carbamazepine-10,11-epoxide 
Chlorothiazide 
Chlorpyrifos 
Clarithromycin 
Climbazole 
Clotrimazole 
Dichlorvos 
Efavirenz 
Emamectin B1a 

Structure
Guanlylurea 
Haloperidol 
Histamine 
Imazalil 
Irgarol 
Ivermectin B1a 
Ketoconazole 
Metazachlor 
Metformin 
Methamidophos 

Structure
Metolachlor 
Nigericin 
Omethoate 
Phenazine 
Progesterone 
Reserpine 
Rifaximin 
Simazine 
Simvastatin 

Structure
Spinosyn A 
Sucralose 
Sudan I 
TCMTB 
Trichlorfon 
Tylosin 
Vancomycin 

Code 1. *The code used for the semi-quantification methods.*

```
library(caret)
library(enviPat)
library(plotly)
library(RRF)
library(rcdk)
library(tidyverse)
library(webchem)

fn_isotope_distribution <- function(smiles){
  molecule <- parse.smiles(smiles)[[1]]
  formula <- get.mol2formula(molecule,charge=0)
  formula <- formula@string

  data(isotopes)
  pattern<-isopattern(isotopes,
                      formula,
                      threshold=0.1,
                      plotit=FALSE,
                      charge=FALSE,
                      algo=1)
  isotopes <- as.data.frame(pattern[[1]])
  isotope_dist <- as.numeric(sum(isotopes$abundance))
  return(isotope_dist)
}

fn_viscosity <- function(organic,organic_modifier){
  viscosity <- case_when(
    organic_modifier == "MeCN" ~ (-
0.000103849885417527)*organic^2+0.00435719229180079*organic+0.884232
851261593,
    organic_modifier == "MeOH" ~ (-
0.00035908)*organic^2+0.031972067*organic+0.90273943)
  return(viscosity)
}

fn_surface_tension <- function(organic,organic_modifier){
  surface_tension <- case_when(
    organic_modifier == "MeCN" ~ 71.76-
2.906*71.76*(organic/100)+(7.138*27.86+2.906*71.76-
71.76)*(organic/100)^2+(27.86-7.138*27.86)*(organic/100)^3,
    organic_modifier == "MeOH" ~ 71.76-
2.245*71.76*(organic/100)+(5.625*22.12+2.245*71.76-
71.76)*(organic/100)^2+(22.12-5.625*22.12)*(organic/100)^3)
  return(surface_tension)
}

fn_polarity_index <- function(organic,organic_modifier){
  polarity_index <- case_when(
    organic_modifier == "MeCN" ~ (organic/100)*5.1+((100-
organic)/100)*10.2,
    organic_modifier == "MeOH" ~ (organic/100)*5.1+((100-
organic)/100)*10.2)
  return(polarity_index)
}
```

```
fn_organic_percentage <- function(eluent_parameters,ret_time){
  ApproxFun <- approxfun(x = eluent_parameters$time, y =
eluent_parameters$B)
  organic <- ApproxFun(ret_time)
  return(organic)
}

assigning_closest_cal_comp_RF <- function(xRT, cal_compounds) {
  RF_cal <- cal_compounds %>% slice(which.min(abs(xRT - ret_time)))
%>%select(RF)
  print(unlist(RF_cal))
}

assigning_closest_cal_comp <- function(xRT, cal_compounds) {
  Comp_cal <- cal_compounds %>% slice(which.min(abs(xRT -
ret_time))) %>%select(Compound)
  print(unlist(Comp_cal))
}

assigning_closest_cal_comp_RT <- function(xRT, cal_compounds) {
  Comp_cal <- cal_compounds %>% slice(which.min(abs(xRT -
ret_time))) %>%select(ret_time)
  print(unlist(Comp_cal))
}

SB1 <- read_delim('Quantem_SB1_w_compname.csv',
                  delim = ';',
                  col_names = TRUE)

SB1 <- SB1 %>%
  group_by(SMILES) %>%
  mutate(IC = fn_isotope_distribution(SMILES)) %>%
  ungroup()


SB1 <- SB1 %>%
  mutate(signal = signal*IC,

         RF = signal / concentration)

SA <- SB1 %>%
  filter(`SB/SA comp.` == "SA")

similarity_matrix <- read_delim('similarity_matrix.csv',
                                delim = ";",
                                col_names = TRUE)

similarity_matrix <- similarity_matrix %>%

  rename(SB_compound = Compound) %>%
  rename(Compound = Similar) %>%
  group_by(SB_compound) %>%
  slice(which.max(similarity_per_cent)) %>%
  ungroup() %>%
  left_join(SA) %>%
```

```
  select(SB_compound, Compound, RF, ret_time, similarity_per_cent,
Maximum_expected_error) %>%
  rename(Similar = Compound) %>%
  rename(Compound = SB_compound,
         RF_similar_compound = RF,
         ret_time_similar_compound = ret_time)

SB1 <- SB1 %>%
  left_join(similarity_matrix) %>%
  mutate(c_pred_similar_compound = signal / RF_similar_compound)

cal_Comp_closeRT <- c()
for(i in 1:69){
 cal_Comp_closeRT <- c(cal_Comp_closeRT,
assigning_closest_cal_comp(SB1[i,]$ret_time, SA))
}

cal_RF_closeRT <- c()
for(i in 1:69){
  cal_RF_closeRT <- c(cal_RF_closeRT,
assigning_closest_cal_comp_RF(SB1[i,]$ret_time, SA))
}

cal_RT_closeRT <- c()
for(i in 1:69){
  cal_RT_closeRT <- c(cal_RT_closeRT,
assigning_closest_cal_comp_RT(SB1[i,]$ret_time, SA))
}

SB1 <- data.frame(SB1, cal_Comp_closeRT, cal_RF_closeRT,
cal_RT_closeRT)

SB1 <- SB1 %>%
  mutate(c_pred_close_ret_time = signal / cal_RF_closeRT)

regressor_pos <- read_rds("ESIpos_model_191116.rds")
descs_pos <-  read_rds("ESIpos_model_descs_191116.rds")

Padel_data <-  read_delim('descs200629.csv',
                          delim = ",",
                          col_names = TRUE,
                          trim_ws = TRUE)

eluent_parameters <- read_delim('eluent.csv',
                                delim = ";",
                                col_names = TRUE)

organic_modifier <- "MeCN"
pH <- 2.7
NH4 <- 0

SB1 <- SB1 %>%
  mutate(
    organic_modifier = organic_modifier,
    organic = fn_organic_percentage(eluent_parameters, ret_time),
    pH.aq. = pH,
```

```
    NH4 = NH4,
    viscosity =  fn_viscosity(organic,organic_modifier),
    surface_tension = fn_surface_tension(organic,organic_modifier),
    polarity_index = fn_polarity_index(organic,organic_modifier))

prediction_set_model_pos <- SB1 %>%
  left_join(Padel_data) %>%
  select(Compound, SMILES, descs_pos) %>%
  na.omit() %>%
  mutate(logIE_pred = 0)
prediction <-  predict(regressor_pos, newdata =
prediction_set_model_pos, predict.all = TRUE)
prediction <- prediction$aggregate
prediction_set_model_pos <- prediction_set_model_pos %>%
  mutate(logIE_pred = prediction) %>%
  select(Compound, SMILES,logIE_pred)

SB1 <- SB1 %>%
  left_join(prediction_set_model_pos)

lin_fit_logRF <- lm(log(RF, 10) ~ logIE_pred,
                    data = SB1 %>% filter(SB.SA.comp. == "SA")
                    %>% filter(Compound != "Butocarboxim [M+NH4]+" &
                               Compound != "Clarithromycin [M+Na]+"
&
                               Compound != "Ivermectin [M+NH4]+" &
                               Compound != "Nigericin [M+NH4]+" &
                               Compound != "Simvastatin [M+Na]+" &
                               Compound != "Simvastatin [M+NH4]+" &
                               Compound != "Sucralose [M+Na]+" &
                               Compound != "Sucralose [M+NH4]+"))

SB1 <- SB1 %>%
  mutate(logRF_pred = lin_fit_logRF$coefficients[2]*logIE_pred +
lin_fit_logRF$coefficients[1],
         c_pred_IE = signal/(10^logRF_pred))

similarity_matrix <- read_delim('similarity_matrix.csv',
                                delim = ";",
                                col_names = TRUE)

similarity_matrix <- similarity_matrix %>%
  rename(SB_compound = Compound) %>%
  rename(Compound = Parent_compound) %>%
  group_by(SB_compound) %>%
  slice(which.max(similarity_per_cent)) %>%
  ungroup() %>%
  left_join(SA) %>%
  select(SB_compound, Compound, RF, ret_time, similarity_per_cent,
Maximum_expected_error) %>%
  rename(Parent_compound = Compound) %>%
  rename(Compound = SB_compound,
         RF_parent_compound = RF,
         ret_time_parent_compound = ret_time)

SB1 <- SB1 %>%
```

```
  left_join(similarity_matrix) %>%
  mutate(c_pred_parent_compound = signal / RF_parent_compound)


SB1 <- SB1%>%
  mutate(similar_c_comparison = case_when(concentration <
c_pred_similar_compound ~ c_pred_similar_compound / concentration,
                                          concentration >
c_pred_similar_compound ~ concentration / c_pred_similar_compound),
         closeRT_c_comparison = case_when(concentration <
c_pred_close_ret_time ~ c_pred_close_ret_time / concentration,
                                          concentration >
c_pred_close_ret_time ~ concentration / c_pred_close_ret_time),
         IE_c_comparison = case_when(concentration < c_pred_IE ~
c_pred_IE / concentration,
                                     concentration > c_pred_IE ~
concentration / c_pred_IE),
         parent_c_comparison = case_when(concentration <
c_pred_parent_compound ~ c_pred_parent_compound / concentration,
                                         concentration >
c_pred_parent_compound ~ concentration / c_pred_parent_compound))

all_c_comparison_SB1 <- SB1%>%
  select(Compound, concentration, SB.SA.comp.,
c_pred_similar_compound,
         c_pred_close_ret_time, c_pred_IE, c_pred_parent_compound,
         similar_c_comparison, closeRT_c_comparison,
IE_c_comparison,
         parent_c_comparison, ret_time, RF, similarity_per_cent)

all_c_comparison_SB1_error <- gather(data = all_c_comparison_SB1,
                                     key = "SQ_approach",
                                     value = error,
similar_c_comparison,
                                     closeRT_c_comparison,
IE_c_comparison,
                                     parent_c_comparison) %>%
  select(Compound, concentration, SB.SA.comp., ret_time, RF, error,
SQ_approach, similarity_per_cent)

all_c_comparison_SB1_pred_c <- gather(data = all_c_comparison_SB1,
                                      key = "SQ_approach",
                                      value = pred_conc,
c_pred_similar_compound,
                                      c_pred_close_ret_time,
c_pred_IE, c_pred_parent_compound, similarity_per_cent) %>%
  mutate(SQ_approach = case_when(
    SQ_approach == "c_pred_similar_compound" ~
"similar_c_comparison",
    SQ_approach == "c_pred_close_ret_time" ~ "closeRT_c_comparison",
    SQ_approach == "c_pred_IE" ~ "IE_c_comparison",
    SQ_approach == "c_pred_parent_compound" ~ "parent_c_comparison"
  )) %>%
  select(Compound, concentration, SB.SA.comp., ret_time, RF,
pred_conc, SQ_approach)
```

```
all_c_comparison_SB1 <- all_c_comparison_SB1_error %>%
  left_join(all_c_comparison_SB1_pred_c)

all_c_comparison_SB1 <- all_c_comparison_SB1 %>%
  filter(SB.SA.comp. == "SB")%>%
  filter(error !="NA")

all_c_comparison_SB1 <- all_c_comparison_SB1 %>%
  mutate(sample = "SB1")

all_c_comparison_SB1 <- all_c_comparison_SB1 %>%
  select(Compound, concentration, ret_time, RF, similarity_per_cent,
SQ_approach, error,
         pred_conc, sample)

error_stat_SB1 = all_c_comparison_SB1 %>%
  group_by(SQ_approach) %>%
  summarise(
    mean_error_c = mean(error),
    median_error_c = median(error),
    max_error_c = max(error),
    quantile_error_c = quantile(error, probs = c(0.95)),
    n_dp = length(error),
    n_dp_less_than_ten = length(error[error<10 & error>1]),
    percentage_less_than_ten = (length(error[error<10 &
error>1]))/(length(error)))%>%
  mutate(sample = "SB1") %>%
  ungroup()
```